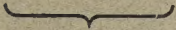


Série A, N° 3392

N° d'ordre

4264



THÈSES

PRÉSENTÉES

A LA FACULTÉ DES SCIENCES
DE L'UNIVERSITÉ DE PARIS

POUR OBTENIR

LE GRADE DE DOCTEUR ÈS SCIENCES MATHÉMATIQUES

PAR

Pierre THIONET

PREMIÈRE THÈSE

La perte d'information par sondage.
(application de la notion d'information à la théorie
de l'échantillonnage)

DEUXIÈME THÈSE

Propositions données par la Faculté

Soutenues le 16 Juin 1958 devant la Commission d'examen

MM. G. DARMOIS

PRÉSIDENT

DUGUÉ

VILLE

}

EXAMINATEURS

Série A, N° 3392

N° d'ordre

4264



THÈSES

PRÉSENTÉES

A LA FACULTÉ DES SCIENCES
DE L'UNIVERSITÉ DE PARIS

POUR OBTENIR

LE GRADE DE DOCTEUR ÈS SCIENCES MATHÉMATIQUES

PAR

Pierre THIONET

PREMIÈRE THÈSE

La perte d'information par sondage.
(application de la notion d'information à la théorie
de l'échantillonnage)

DEUXIÈME THÈSE

Propositions données par la Faculté

Soutenues le 16 Juin 1958 devant la Commission d'examen

MM. G. DARMOIS

PRÉSIDENT

DUGUÉ

VILLE

}

EXAMINATEURS

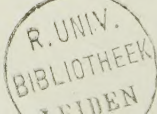
FACULTÉ DES SCIENCES DE L'UNIVERSITÉ DE PARIS

Doyen J. PÉRÉS

- PROFESSEURS -

G. JULIA	T Analyse sup. & Alg. sup.	FELDMANN	T Biologie (P.C.B.).
FOCH	T Méc. physique & experim.	GUINIER	T Physique des solides (P.C.B.).
PAUTHENIER	T Electrotechnique Cle.	JOST	T Physiologie comparée
DE BROGLIE	T Théories physiques.	FORTET	T Calc. des probabilités.
FRENANT	T Anatomie et Histologie comparées.	SCHWARTZ	T Calc. diff. et intégral.
GARNIER	T Géométrie supérieure.	CHOUARD	T Physiologie Végétale.
PERES	T Mécan. des fluides et applications.	MALAVARD	T Aviation.
LAUGIER	T Physiologie générale.	BRELOT	T Calcul différentiel.
M. CURIE	T Physique (S.P.C.N.).	NORMANT	T Chimie organique (P.C.B.).
BARRABE	T Géol. struct. & Géol. appliquée.	BENARD	T Chimie (P.C.B.).
G. DARMOIS	T Calcul des probabil. et physique Math.	BUVAT	T Botanique (E.N.S.).
J. BOURCART	T Géogr. phys. & Géol. dynamique.	DUGUE	T Statistiques mathém.
PLANTEFOL	T Botanique.	GELOSO	T Chimie (S.P.C.N.).
GRASSÉ	T Evol. des êtres organ.	SOULAIRAC	T Psychophysiologie.
PREVOST	T Chimie organique.	ULRICH	T Physiol. Végétale.
BOULIGAND	T Appl. de l'analyse à la Géométrie.	MARECHAL	T Optique théorique et appliquée.
CHAUDRON	T Chimie appliquée.	KIRRMANN	T Théories chimiques.
WYART	T Minéralogie.	CHADEFAUD	T Botanique.
TEISSIER	T Zoologie.	DESTRIAU	T Physique (M.P.C.).
MANGENOT	T Biologie végétale.	Mlle LE BRETON	T Physiologie générale.
P. AUGER	T Physique quantique et Relativité.	SALEM	T Mathém. générales.
MONNIER	T Physiol. générale.	LELONG	T Mathém. générales
FIVETEAU	T Paléontologie.	DEVILLERS	T Anatomie et Histologie comparées.
ROCARD	T Physique (E.N.S.).	EHRESMANN	T Topologie Algébrique.
H. CARTAN	T Mathématiques (E.N.S.).	FRANÇON	T Physique (S.P.C.N.).
LAFFITTE	T Chimie Générale.	GLANGEAUD	T Géogr. physique et Géol. dynamique.
FAVARD	T Géométrie supér.	GODEMENT	T Mathématiques (M.P.C.).
COULOMB	T Physique du Globe.	PISOT	T Calcul des Probabilités.
Mlle COUSIN	T Biologie animale (S.P.C.N.).	ROCH (E.)	T Géologie.
CHRETIEN	T Chimie minérale.	SCHATZMANN	T Astrophysique.
P. DRACH	T Zoologie.	TERMIER	T Paléontologie stratigr.
KASTLER	T Physique.	ZAMANSKY	T Mathém. générales
EPHRUSSI	T Génétique.	LENNUIER	T Physique (P.C.B.).
WURMSER	T Biol. physico-chimique.	RIZET	T Génétique.
RIVIERE	T Géologie (S.P.C.N.).	ROUTHIER	T Géologie Appliquée.
GAUTHERET	T Biol. Végétale (P.C.B.).	Mme TONNELAT	T Théories physiques.
LUCAS	T Recherches physiques.	DIXMIER	T Mathématiques (M.P.C.).
A. THOMAS	T Biologie cellulaire.	SOUCHAY	T Chimie générale.
ARNULF	T Optique appliquée.	AGRAIN	T Physique (P.C.B.).
M. MORAND	T Enseignement de phys.	BRUSSET	T Chimie Minérale.
SOLEILLET	T Physique.	M. LEVY	T Physique théorique.
FORTIER	T Mécanique exper. des fluides.	LENORMANT	T Physiologie.
DANJON	T Astronomie.	Mme CHAIX	T Chimie biologique.
LAPORTE	T Phys. générale & Radioactivité.	Mme HUREL-PY	T Biol. Végétale (P.C.B.).
JANET	T Méc. anal. et Mécanique céleste.	PIAUX	T Chimie (P.C.B.).
PETIT	T Biol. maritime.	BRUN	T Méc. exper. des fluides.
QUENEY	T Météorologie et dyn. atmosphérique.	LEDERER	T Chimie biologique.
GALLIEN	T Embryologie.	Mme DUBREIL	T Mathématiques.
EICHHORN	T Biol. Vég. (S.P.C.N.).	Mlle LELONG	T Mathématiques (E.N.S.).
DE CUGNAC	T Biol. Vég.	BELLAIR	T Géologie (S.P.C.N.).
Mlle CAUCHOIS	T Chimie Physique.	COTTE	T Physique (M.P.C.).
THELLIER	T Physique du Globe.	DUBOIS J. E.	T Chimie (S.P.C.N.).
L'HERITIER	T Génétique.	LAMOTTE	T Zoologie (E.N.S.).
GRIVET	T Radioélectricité.	OLMER	T Chimie (S.P.C.N.).
PONCIN	T Méc. générale.	ROUAULT	T Energétique générale.
DUBREIL	T Arith. & théorie des nombres.	GAUTHIER	T Physique générale.
QUELET	T Chimie organique.	BARCHEWITZ	T Mécanique appliquée.
CAGNIARD	T Géophysique appliquée.	BROSSEL	T Chimie Physique.
CHAMPETIER	T Chimie macromoléc.	BUSER	T Physique Atomique.
CUVILLIER	T Géol. struct. et Géol. appliquée.	CAMUS	T Biologie Animale (P.C.B.).
JUNG	T Pétrographie.	Castaing	T Biologie Végétale (P.C.B.).
TRILLAT	T Microscopie et diff. électronique.	CURIEN	T Physique Générale.
WIEMANN	T Chimie organique et structurale.	MOYSE	T Minéralogie et Cristallog.
JACQUINOT	T Spectroscopie et Phys. céleste.	NOIROT	T Physiologie Végétale.
VASSY	T Phys. de l'atmosphère.	PANNETIER	T Evolution des Etres organ.
DESTOUCHES	T Théories physiques.	POSSOMPES	T Chimie (P.C.B.).
M. PRUVOST	T Géologie.	PULLMANN	T Biologie Animale (P.C.B.).
AMEL	T Chimie générale (P.C.B.).	TEILLAC	T Chimie théorique.
HOCART	T Minéralogie.	TONNELAT	T Physique Nucléaire et Radioactivité.
J.P. MATHIEU	T Physique (P.C.B.).	VILLE	T Biologie physico-chimique
COUTEAUX	T Biol. animale (P.C.B.).	WILLEMART	T Econométrie
MAY	T Biol. animale (P.C.B.).	DODE	T Chimie (P.C.B.).
CHOQUET	T Calc. différentiel et Intégral.	FREY MANN	T Chimie
		GUINCHET	T Physique (S.P.C.N.).
		ROLLET	T Biologie Végétale (Orsay).
		Mlle JOSIE	T Chimie (P.C.B.).
		CHEVALLEY	T Chimie (P.C.B.).
		DESNUELLE	T Géométrie Algébrique et théorie des groupes.
			T Chimie biologique.

Secrétaire général. Ch. MONIER



Entreprises à partir d'une ample connaissance théorique et pratique de la technique, ces recherches n'auraient pu être menées à bien si nous n'avions été puissamment aidé et encouragé. Monsieur Georges Darmois nous initia en cette Faculté même à la Statistique Mathématique et nous enseigna aussi la géométrie à l'Ecole Normale; bien souvent depuis, sa bonté et sa compréhension nous ont épaulé en des circonstances fort diverses. Avec Monsieur Darmois qui nous conseilla de chercher "du côté de l'information" et a bien voulu nous faire l'honneur de présider la soutenance de cette thèse, nous voulons remercier Monsieur le Professeur Dugué de nous avoir, avec tant de patience, encouragé à la poursuivre jusqu'au bout.

Ce travail demandait toute la compréhension et l'amitié de Monsieur Closon, Directeur Général de l'Institut National des Statistiques et de Monsieur Gruson, Chef du Service des Etudes Economiques et Financières au Ministère des Finances, qui voudront bien trouver ici l'expression de nos sentiments dévoués et reconnaissants.

Il est juste d'associer à nos remerciements Monsieur M. Fréchet (qui eut en 1941 un rôle important dans notre orientation); Monsieur Jean Mayer (Professeur au Collège de France récemment décédé), grâce à qui nous découvrîmes les sondages dès 1945; nos premiers Chefs à la Statistique Générale MM. Bunle et R. Rivet; enfin M. Alfred Sauvy, les Professeurs George Barnard et Hans Kellner, MM. André Nataf, Fonsagrive, Desabie, dont les avis et encouragements nous furent précieux; nous oublions certainement de citer tous ceux qui le méritent et nous en excusons d'avance.

Cette thèse est dédiée à la mémoire de notre cousine Simone Léard et de nos grand-père, tante et grand-oncle - qui furent membres de l'Enseignement des 1er et 2ème degrés.



Digitized by the Internet Archive
in 2026

Pierre THIONET

**LA PERTE D'INFORMATION
PAR SONDAGE**

(Calcul des probabilités)

INTRODUCTION

L'objet de la présente thèse est l'introduction du concept d'information au sens de Schutzenberger dans la théorie des sondages. On est parti de l'idée de la validité, dans ses grandes lignes, de la technique des sondages (telle qu'elle est enseignée dans les manuels de Yates - traduit en français sous la direction de M.G. Darmois, - de Deming, de Hansen, Hurvitz et Madow, de Sukhatme, de Cochran, etc., dans les cours de Tang, Kellerer, Žarcović, etc. Lévy-Bruhl, Chartier, Desabie, et nous-même). On a admis toutefois que les concepts théoriques qui sont à la base de ces méthodes pouvaient gagner à un examen nouveau, - et que la technique s'en trouverait consolidée.

Malgré l'accord qui règne entre tous ces auteurs, il existait en fait deux écoles : celle de Rothamsted pour laquelle les échantillons et populations sont extraits d'univers d'effectifs infinis, où les variables ont des distributions volontiers continues et dotées de densités; celle du Bureau of the CENSUS pour laquelle les échantillons sont extraits de populations-univers d'effectifs finis (ce qui conduit notamment à considérer un coefficient de corrélation intraclasses compris pour les uns entre 1 et 0, et pour les autres entre 1 et $-1/\sqrt{v} - 1$). Nous nous plaçons d'emblée dans l'école américaine, sous réserve qu'on n'oublie pas le cas limite (essentiel en pratique) où v devient infiniment grand.

L'existence de lois de distribution dont le moment d'ordre 2 n'existe pas ne posera pas de problème bien grave (car toutes les informations ne sont définies qu'à un facteur près).

On nous reprochera sans doute davantage d'avoir défini la perte d'information à partir d'un univers complètement connu dont les éléments échappent un à un à notre vue (contrairement aux méthodes d'analyse progressive), au lieu de définir le gain d'information à partir d'un état initial de la connaissance, comme le fait Lindley (suivant de près la définition que Shannon donne de l'information dans la théorie des communications). Lindley n'obtient de résultats qu'en supposant connues et la loi de distribution (école de Rothamsted) et la probabilité a

priori des paramètres à estimer, considérés comme des variables aléatoires (méthode d'estimation d'après Bayes). Cette théorie ne paraît pas, au moins sous sa forme actuelle, applicable aux techniques de sondage (Ch.VIII). Par exemple une quantité d'information qui devient infinie avec le nombre de tirages (avec remise) est inapplicable aux sondages à plusieurs degrés, stratifiés et en grappe.

Au contraire nous avons retrouvé, aux premières places des pertes d'information, la variance d'échantillonnage, dont fait grand usage la technique. Nous avons d'ailleurs critiqué, pour les estimateurs biaisés, le procédé courant qui consiste à ajouter à la variance le carré du biais; ce total n'est plus une perte d'information. Nous avons également signalé que l'usage de la variance n'était pas approprié au cas de l'estimation d'une médiane.

De toute façon la variance est une perte d'information particulière, à savoir symétrique; elle suppose qu'on attribue le même poids à deux erreurs d'échantillonnage, l'une positive, l'autre négative, de même ampleur. - On en a trouvé bien d'autres.

Les expressions de la perte d'information dans le cas de distributions quelconques (Ch.IV) n'ont été trouvées que sous des conditions de régularité très restrictives; il est possible que ces conditions puissent être réduites; en outre il y aurait lieu d'étudier plus systématiquement, les possibilités que peut présenter le cas où la dérivée première n'existe pas quand la perte élémentaire s'annule (ou d'autres cas encore).

Nous n'avons pas été en mesure, d'ailleurs, d'interpréter toute perte d'information au moyen d'estimateurs et de quantités à estimer; il semble que ce soit impossible. - Une autre lacune dans nos recherches serait une étude de l'estimation du mode de l'univers par celui de l'échantillon; il y aurait un réel intérêt pratique à savoir exprimer (et minimiser) la perte d'information correspondante.

Pour l'estimation de la plus grande valeur ou de l'étendue (range), la répartition optimum de l'échantillon sera celle qui donne la moindre perte d'information, par une extension naturelle de la théorie de la moindre variance. Le total [coût d'enquête, plus perte d'information] est à minimiser, ce qui est assez dans les idées de Wald (1950); le multiplicateur de Lagrange joue le rôle d'unité monétaire.

Chemin faisant, nous avons complété la théorie de l'estimation par sondage dans le cas biaisé. On aurait pu croire qu'alors toute statistique pouvait servir à estimer tout paramètre. Du moment qu'on définit l'information par une récurrence entre estimations sur une suite de plans de sondage de tailles décroissantes, l'estimation Z' n'informe sur

ζ que si elle informe sur tous les Z qui s'intercalent entre Z' et ζ . La perte d'information ne peut être définie que pour les estimateurs uniformément biaisés; ce qui signifie qu'il existe une fonctionnelle $\Delta[Z'Z]$ telle qu'on ait $\mathcal{E}\Delta = 0$.

Ceci conduit à calculer la perte d'information dans le cas où il existe une fonctionnelle ψ telle que :

$$\mathcal{E} \psi[Z] = \psi[\zeta]$$

En particulierisant encore, on a introduit l'estimateur isomorphe qui permet de retrouver les cas pratiques courants d'estimation biaisée; il correspond au cas où $\zeta = f(a, b, c)$ $Z = f(A, B, C)$, la fonction f étant la même, A, B, C étant des aléatoires calculés sur l'échantillon, leurs espérances mathématiques étant a, b, c . Les indicateurs que calculent les techniciens ont été retrouvés, par exemple pour les ratios; ce qu'on considérerait jusqu'ici comme des valeurs approchées des carrés des coefficients de variation sont de véritables pertes d'information.

A cette même occasion on a retrouvé l'information de Fisher, comme perte d'information de l'estimation du maximum de vraisemblance - Ch. IV).

Les plans de sondage réels étant notablement plus compliqués que l'urne de Bernoulli, on a mené séparément et parallèlement l'étude de ces deux cas (Ch. I, II). On a eu l'idée de représenter les sondages dans l'urne unique par des points d'un axe et les sondages plus généraux par les points d'un treillis, les distances étant comptées sur les droites du dit treillis (Ch. III). Cet éclatement des difficultés suivant leur nature, topologique ou analytique, s'est révélé fructueux. - La représentation métrique des plans de sondage dans un espace euclidien à un nombre donné de dimensions a paru facile pour les sondages à plusieurs degrés et même à plusieurs phases. - Notre compréhension des formules de variance (apparemment compliquées) s'en est trouvée améliorée (Ch. V, VI).

La perte d'information étant supposée prendre place dans l'enseignement des Sondages, doit-on continuer à parler des informations supplémentaires dont l'emploi permet d'améliorer les estimateurs ou les plans de sondage? Autrement dit, que devient l'information quand on compare plusieurs itinéraires ζ, Z , au lieu de se déplacer le long d'un seul itinéraire? Tel est l'objet du Chapitre VII.

Au total nous pensons que les statisticiens ne devraient pas dire d'un plan de sondage qu'il apporte plus d'information qu'un autre, mais qu'il fait perdre moins d'information que l'autre. Une perte négative

est un gain, par rapport à un sondage de référence. Les diverses théories définissent des niveaux d'information; pour nous le niveau le plus bas correspond à un sondage ($\bar{\omega}$) de taille minimum; pour d'autres il correspond à des renseignements antérieurs permettant d'écrire une loi de probabilité a priori. Quant au niveau le plus élevé, pour nous il est nécessairement à distance finie; d'autres l'autorisent à s'éloigner à l'infini. - Cela mis à part, l'équation de récurrence est la même; la nature des solutions diffère; car le symbole Espérance Mathématique n'a pas du tout le même sens chez Lindley qu'ici même. La topologie serait la même, si nous n'avions trouvé utile de faire jouer à l'infini un rôle spécial : il représente l'absence de toute information utilisable; et il est conforme aux habitudes des techniciens de penser qu'on n'a pas d'information utilisable tant qu'il reste une strate non sondée (Ch. VIII).

On peut penser finalement que la technique classique des sondages (y compris l'estimation du maximum de vraisemblance de Fisher, sort consolidée de notre analyse.

CHAPITRE I

TIRAGES ÉQUIPROBABLES DE BOULES DANS UNE URNE

I - INTRODUCTION -

1/ - Point de départ concret.

Les plans de sondage sont établis en pratique pour estimer certaines grandeurs physiques (économiques, sociales, médicales, etc.) au moyen d'observations limitées à un échantillon désigné par des tirages au sort plus ou moins compliqués.

Dans la plupart des cas, ces grandeurs sont des valeurs totales (superficie cultivée en jute, nombre de personnes travaillant moins de quarante heures par semaine, etc.) ou (ce qui revient au même pour la statistique pratique) des valeurs moyennes ou des "pourcentages" par unité de sondage. On "extrapole", c'est-à-dire qu'on passe des pourcentages ou valeurs moyennes à l'estimation du total, en multipliant par un facteur connu.

Ainsi la technique des sondages est axée sur les problèmes d'estimation de moyennes - compte tenu d'un plan de sondage plus ou moins compliqué.

L'estimation des ratios est un autre problème pratique très important qui ne pourra être étudié dans les premiers chapitres. On la trouvera au chapitre IV.

On commence par l'étude du plan de sondage le plus simple : tirages équiprobables de boules dans une urne, sans remise puis avec remise.

2/ - Définitions.

Considérons une urne (ou univers) U renfermant v boules numérotées $1, 2, \dots, i \dots, v$ portant chacune un nombre réel x_i . Désignons par X l'ensemble de ces v valeurs.

La moyenne des x_i sera notée :

$$\sum_i x_i / v = \bar{x}$$

L'écart-type des x_i est σ , défini par :

$$\sum_i (x_i - \bar{x})^2 / v = \sigma^2$$

On tire au sort une boule de l'urne, n fois de suite sans jamais remettre les boules tirées. Pour un échantillon donné de n boules, la moyenne des n valeurs de x_i ainsi connues sera notée :

$$S_i x_i / n = \bar{X}_n \text{ ou } \bar{X}$$

$\bar{X}_n$ est une variable aléatoire lorsque (n étant donné) on envisage les m échantillons également probables à extraire de l'urne, avec chacun desquels on peut calculer une valeur de $\bar{X}_n$.

Cas des tirages sans remise ou exhaustifs .. $m = v! / n! (v - n)!$

Cas des tirages avec remise ou bernoulliens $m = v^n$

Lorsque $\bar{x}$ n'est pas connu, on l'"estime" par une certaine valeur de $\bar{X}_n$, à savoir celle calculée sur l'échantillon effectivement tiré.

On dit que $\bar{X}_n$ est estimateur "sans biais" de x , parce qu'on a :

$$\mathcal{E} \bar{X}_n = \bar{x}$$

où l'opérateur $\mathcal{E}$ désigne l'espérance mathématique (étendue aux m valeurs également probables de $\bar{X}_n$).

Ce qu'on dit pour $\bar{x}$ et $\bar{X}$ s'étend aux moments d'ordre quelconque de X calculés sur l'univers et sur échantillon (avec la même origine pour X).

Plus généralement on considèrera l'estimateur sans biais Z de l'estimé ζ définis sur l'échantillon et sur l'univers respectivement, avec $\mathcal{E} Z = \zeta$.

Erreur d'échantillonnage :

On désigne ainsi l'aléatoire $\bar{X} - \bar{x}$, ou encore $Z - \zeta$.

On a coutume de repérer l'ampleur des erreurs d'échantillonnage par la variance de l'estimateur, $\forall Z = \mathcal{E}(Z - \zeta)^2$. On sait que :

$$\mathcal{V} \bar{X}_n \begin{cases} = (\sigma^2/n) (v - n/v - 1) & (\text{si tirages exhaustifs}) \\ = \sigma^2/n & (\text{si tirages bernoulliens}) \end{cases}$$

On emploie également beaucoup la racine carrée de cette variance :

$$\sigma(Z) = \sqrt{\mathcal{V}(Z)}$$

ou écart-type de Z , et la valeur relative de cet écart-type

$$\sigma(Z)/\zeta = C.V(Z) = \gamma$$

ou coefficient de variation (souvent exprimé en pour cent).

II - LEMMES -

Lemme 1.

Le carré de l'écart-type de v valeurs de (x) peut s'écrire :

$$\sigma^2 = \frac{\sum \delta^2}{v^2}$$

δ^2 désignant le carré de la différence de deux quelconques des valeurs de x ,

$\sum$ désignant la somme des $\frac{v(v-1)}{2}$ valeurs distinctes de δ^2 .

Autre forme du Lemme : $\frac{v\sigma^2}{v-1} = \frac{\Delta^2}{2}$

où $\Delta^2 = \sum \delta^2 / \frac{v(v-1)}{2}$ désigne la moyenne des écarts carrés.

Démonstration du Lemme.

$$\begin{aligned} \delta^2 &= \sum_{i=1}^v \left[(v-1)x_i - \sum_{j \neq i} x_j \right]^2 / v^3 \\ &= \sum_i \left[\sum_{j \neq i} (x_i - x_j) \right]^2 / v^3 \\ &= \left[\sum_{i=1}^v \sum_{j \neq i} (x_i - x_j)^2 + \sum_{i=1}^v \sum_{\substack{j \\ j \neq h}} \sum_h (x_i - x_j)(x_i - x_h) \right] / v^3 \end{aligned}$$

Les carrés distincts $(x_i - x_j)^2$ sont au nombre de $v(v-1)/2$

$\sum_{j \neq i} (x_i - x_j)^2$ comprend $(v-1)$ termes

$\sum_{i=1}^v \sum_{j \neq i} (x_i - x_j)^2$ comprend $(v - 1)$ termes

$\sum_{\substack{j \\ j \neq i \\ i \neq h}} \sum_{\substack{h \\ h \neq j \\ h \neq i}} (x_i - x_j)(x_i - x_h)$ comprend $(v - 1)(v - 2)$ produits.

En sommant cette dernière expression par rapport à i , on obtient $(v - 1)(v - 2)$ produits.

Or les $(x_i - x_j)(x_i - x_h)$ comprennent $\frac{v(v - 1)(v - 2)}{2}$ valeurs distinctes, de sorte que

$\sum_i \sum_j \sum_h (x_i - x_j)(x_i - x_h)$ est égal à deux fois la somme de tous les produits distincts.

Mais :

$$\begin{aligned} (x_i - x_j)(x_i - x_h) + (x_j - x_h)(x_j - x_i) &= (x_i - x_j)(x_i - x_h + x_h - x_j) \\ &= (x_i - x_j)^2 \end{aligned}$$

Les $(v - 1)(v - 2)$ produits donnent donc naissance à $(v - 1)(v - 2)/2$ carrés, c'est-à-dire $(v - 2)$ fois chaque carré distinct, c'est-à-dire

$$(v - 2) \sum \delta^2$$

Ainsi :

$$\begin{aligned} v^3 \delta^2 &= 2 \sum \delta^2 + (v - 2) \sum \delta^2 \\ &= v \sum \delta^2 \end{aligned}$$

d'où

$$\sigma^2 = \sum \delta^2 / v^2 \quad \text{c. q. f. d.}$$

En particulier si la variable x prend les valeurs a et b , le carré de l'écart-type de x peut s'écrire :

$$\begin{aligned} \sigma^2 &= \left[\left(a - \frac{a+b}{2} \right)^2 + \left(b - \frac{a+b}{2} \right)^2 \right] / 2 \\ &= (a - b)^2 / 4 \end{aligned}$$

Lemme 2 sur la variance d'un échantillon exhaustif.

Pour l'échantillon d'effectif n , la moyenne des différences carrées sans répétition est :

$$D = 2 S(x_i - x_j)^2 / n(n-1)$$

où S désigne la sommation étendue à l'échantillon. Il est évident que

$$1/ \quad E(D) = \Delta$$

où E désigne l'espérance mathématique lorsque n est donnée.

$$2/ \quad \frac{n-1}{2n} D = s^2 = S_i \frac{(x_i - \bar{X})^2}{n}$$

Il en résulte que :

$$E \left(\frac{ns^2}{(n-1)} \right) = \frac{v\sigma^2}{v-1}$$

Tel est le lemme 2. (Résultat connu, parfois oublié).

Lemme 3 sur la variance d'un échantillon bernoullien.

Faisons tendre v vers l'infini; il vient à la limite :

Lemme 3 :

$$E' \left(\frac{ns'^2}{n-1} \right) = \sigma^2$$

E' désignant l'espérance mathématique relative aux tirages avec remise. (Résultat bien connu).

III - ETUDE DES SUITES DE PLANS DE SONDAGE -

1/ - Les deux points de vue :

Au lieu d'envisager, comme d'habitude, des tirages successifs dans l'urne, au besoin jusqu'à épuisement ($n = v$), - on pourra, en sens inverse, partir d'une urne parfaitement connue ($v = n$) c'est-à-dire totalement vide (les boules étant placées dans une urne auxiliaire) et extraire de cette dernière, par tirage au sort, successivement

$$1, 2, \dots (v-n), \dots (v-1), v \text{ boules,}$$

qu'on replongera dans l'urne primitive.

2/ - Suite de plans de sondage : notion d'information.

Cas des sondages exhaustifs.

On peut symboliser la famille de sondages en question par des points consécutifs numérotés $0, 1, 2, \dots, n, \dots, v$ sur un axe D, le point n représentant l'ensemble des sondages exhaustifs d'effectif n (pour la population et la variable X considérées).

Une échelle arithmétique sur l'axe D peut d'abord être employée

$$(D) \quad \begin{array}{ccccccccccc} 0 & 1 & 2 & 3 & & n & & v-1 & v \\ : & : & : & : & : & : & : & : & : \end{array}$$

Si l'on se déplace de (0) vers (v), on connaît de mieux en mieux l'urne; (0) représente l'absence totale de renseignements et (v) figure la connaissance parfaite de l'univers. (On désignera plutôt un tel point par un (U).

On peut dire qu'on "gagne de l'information" en passant de (n) à (n + 1) et qu'on "perd de l'information" en passant de n à (n - 1).

Il serait intéressant que la distance (vn) sur l'axe D représentât la perte d'information résultant du fait qu'on a limité les observations à n boules sur v (en ce qui concerne la connaissance de l'urne U). A cet égard l'échelle arithmétique ne nous paraît pas satisfaisante quantitativement; il ne serait pas conforme à notre concept d'information de dire qu'on perd la même quantité d'information chaque fois que l'effectif de l'échantillon diminue d'une unité. Bien au contraire notre sentiment est que :

- connaître toutes les boules, ou toutes sauf une, n'est pas très différent ;
- connaître une boule, ou n'en connaître aucune, est très différent;
- connaître deux boules au lieu d'une constitue un gain d'information bien plus grand que connaître v boules au lieu de v - 1.

En revanche portons sur un axe D les valeurs pour $n = 0, 1, 2, \dots, v - 1$, de la variance de $\bar{X}_n$

$$v \bar{X}_n = \frac{\sigma^2}{n} \frac{v - n}{v - 1}$$

c'est-à-dire :

$$v \bar{X}_n = \frac{v \sigma^2}{v - 1} \left(\frac{1}{n} - \frac{1}{v} \right)$$

Le point (0) est rejeté à l'infini, tandis que le point (v) est l'origine des abscisses. Quand n croît, chaque pas fait vers la connaissance parfaite est bien inférieur au précédent.



Ainsi notre intuition sera satisfaite si l'on choisit la variance $\mathcal{V} \bar{X}$ pour mesurer la perte d'information imputable au sondage. $\mathcal{V} \bar{X}$ possède une autre propriété.

THEOREME -

La variance $\mathcal{V} \bar{X}$ satisfait à la définition de l'information suivant l'axiome de Schützenberger (Sch. page 45).

En effet, posons :

$$p(v, n) = \mathcal{V} \bar{X}_n$$

symbole rappelant que n unités sur v sont connues; et montrons que :

$$p(v, n-j) - p(v, n) = E P(n, n-j)$$

où P a la même définition que p mais est aléatoire.

Pour cela formons :

$$\begin{aligned} p(v, n-j) - p(v, n) &= \frac{v\sigma^2}{v-1} \left(\frac{1}{n-j} - \frac{1}{v} - \frac{1}{n} + \frac{1}{v} \right) \\ &= \frac{v\sigma^2}{v-1} \left(\frac{1}{n-j} - \frac{1}{n} \right) \end{aligned}$$

et posons :

$$P(n, n-j) = \frac{ns^2}{n-1} \left(\frac{1}{n-j} - \frac{1}{n} \right)$$

où s^2 désigne la variance d'un échantillon (exhaustif) de n boules :

$$s^2 = S(x_i - \bar{X})/n$$

Le théorème découle immédiatement du Lemme 2 :

$$E(ns^2/n-1) = v\sigma^2/v-1$$

3/ - Cas des sondages Bernoulliens.

Le nombre n de tirages peut (théoriquement au moins) être pris aussi grand qu'on veut (et bien entendu dépasser ν).

Représentons par des points consécutifs sur l'axe D, les divers plans de sondage. Lorsque l'effectif n croît

$$n = 1, 2, 3 \dots$$

le point figuratif (n) part d'une position (1) et se rapproche constamment sans jamais y parvenir d'un point U figurant la connaissance parfaite de l'urne.

Le point (n) a donc un point d'accumulation ∞ et il n'est pas évident que celui-ci doive coïncider avec le point U figurant la connaissance parfaite.



Nous ferons par la suite le Postulat $\infty = U$, c'est-à-dire que la perte d'information, par rapport à U, tend vers 0 quand n tend vers l'infini.

4/ - Accord entre figurations des sondages exhaustif et bernoullien.

Superposons les axes D relatifs à un sondage exhaustif et à un sondage bernoullien dans la même population.

Il est évident que les deux points (1) doivent coïncider, puisque sondages bernoullien et exhaustif sont indiscernables tant qu'on ne tire qu'une boule. De même le point (ν) de l'axe "exhaustif" doit coïncider avec le point U de l'axe bernoullien; et, avec le Postulat ($\infty = U$), il coïncide avec le point ∞ .

En outre, il est intuitif que le tirage de la $n^{\text{ème}}$ boule apporte moins d'information en sondage bernoullien qu'en sondage exhaustif, ce qui doit se traduire par :

$$\begin{array}{ccc} \text{distance } n - 1, n & & \text{distance } n - 1, n \\ \text{sur D exhaustif} & \begin{array}{c} > \end{array} & \text{sur D bernoullien} \end{array}$$

Toutes ces conditions sont remplies si la perte d'information en sondage bernoullien est choisie égale à

$$\frac{\sigma^2}{n}$$

quand la perte en sondage exhaustif est elle-même (à effectif égal)

$$\frac{\sigma^2}{n} = \frac{v - n}{v - 1}$$

Examinons toutefois ce qu'il advient de la formule de Schützenberger relative aux quantités d'information.

5/ - THEOREME -

La variance $\mathcal{V} \bar{X}$ en sondage bernoullien satisfait à la définition de l'information suivant l'axiomatique de Schützenberger.

En effet, posons :

$$p'(v, n) = \frac{\sigma^2}{n}$$

il vient

$$p'(v, n - 1) - p'(v, n) = \frac{\sigma^2}{n(n - 1)}$$

Calculons ensuite

$$P'(n, n - 1)$$

Soit s'^2 la variance des x_i de l'échantillon tiré de taille n . Pour passer de là à l'échantillon d'effectif $n - 1$, il faut tirer au sort l'une des n boules et l'exclure, ce qui équivaut à conserver un échantillon exhaustif de $(n - 1)$ boules; la perte d'information correspondante est donc :

$$P'(n, n - 1) = \frac{s'^2}{n - 1} \cdot \frac{n - (n - 1)}{n - 1} = \frac{s'^2}{(n - 1)^2}$$

Il faut enfin prendre l'espérance mathématique de s'^2 , la sommation étant étendue à tous les échantillons bernoulliens d'effectif n . Appliquons la lemme 3, il vient :

$$\mathcal{E}' \frac{s'^2}{(n - 1)^2} = \frac{1}{(n - 1)^2} \times \frac{n - 1}{n} \sigma^2 = \frac{\sigma^2}{n(n - 1)}$$

D'où

$$p'(v, n - 1) - p'(v, n) = \mathcal{E}' P'(n, n - 1) \quad \text{c. q. f. d.}$$

6/ - Extension à la variance de tout estimateur sans biais.

Soit Z_n et $Z_{n'}$, les estimateurs de ζ sur échantillons de tailles n et n'

$$\mathcal{E}(Z_n) = \zeta, \quad \mathcal{E}(Z_{n'}) = \zeta, \quad \underline{n > n'}$$

Posons

$$p(v, n) = \mathcal{E}(Z_n - \zeta)^2$$

$$p(v, n') = \mathcal{E}(Z_{n'} - \zeta)^2$$

$$P(n, n') = \mathcal{E}(Z_{n'} - Z_n)^2$$

où E désigne l'espérance mathématique de $Z_{n'}$ sur les échantillons de taille n' tirés au sort dans un certain échantillon de taille n . Il vient :

$$\begin{aligned} p(v, n') - p(v, n) &= \mathcal{E} E[(Z_{n'} - Z_n)^2] \\ &= \mathcal{E} E\{(Z_{n'} - Z_n)[(Z_{n'} + Z_n - 2\zeta) - (Z_{n'} - Z_n)]\} \\ &= 2 \mathcal{E}[(Z_n - \zeta) E(Z_{n'} - Z_n)] \end{aligned}$$

qui est nul si $Z_{n'}$ est estimateur sans biais de Z_n .

$$E Z_{n'} = Z_n \quad (\text{quel que soit } n, v \geq n \geq n')$$

Cette condition renferme $\mathcal{E} Z_{n'} = \zeta$ et $\mathcal{E} Z_n = \zeta$.

7/ - Définition.

Nous dirons alors que l'estimateur est uniformément sans biais. Alors :

$$\forall Z = \mathcal{E}(Z - \zeta)^2$$

est perte d'information (au sens de l'axiomatique de Schützenberger).

Remarque.

Il n'est pas exclu que la relation de récurrence puisse être vérifiée sous des conditions moins strictes, n et n' ne prenant que certaines valeurs entières par exemple.

Cas des estimations isomorphes.

Si ζ et Z sont des expressions isomorphes⁽¹⁾ (des x_i de l'univers et de l'échantillon), l'absence de biais est uniforme.

(1) C'est-à-dire ont même expression analytique.

Exemples.

$\bar{x}$ et $\bar{X}$ sont uniformément sans biais; de même $v\sigma^2/v-1$ et $nS^2/n-1$

Exemple d'estimateur sans biais.

Si on estime la moyenne d'une distribution symétrique par la médiane de l'échantillon (n et n' impairs), on a bien

$$\mathcal{E} Z_n = \zeta, \quad \mathcal{E} Z_{n'} = \zeta.$$

mais on n'a pas en général

$$E Z_{n'} = Z_n$$

et la variance $\mathcal{V} Z_n$ n'est pas une perte d'information.

Exception.

$X = a, b, c, d$ ou e , avec $\zeta = c$, $\overline{ce} = -\overline{ca}$, $\overline{cd} = -\overline{cb}$, $\overline{ca} = 7\overline{cb}/2$ avec $v = 5$; $n = 3$; $n' = 1$; on a bien

$$\mathcal{E}(Z_3 - \zeta) (E Z_1 - Z_3) = 0$$

d'où

$$\mathcal{V} Z_1 - \mathcal{V} Z_3 = \mathcal{E}(Z_1 - Z_3)^2$$

D'une manière générale on évitera de s'intéresser à des distributions trop particulières qui peuvent avoir des propriétés exceptionnelles sans portée pratique.

8/ - Cas de l'Estimation de la variance.

Partant de la définition de σ^2 , du Lemme 1

$$\sigma^2 = \sum_i \frac{(x_i - x_j)^2}{v^2}$$

ou mieux :

$$\frac{v\sigma^2}{v-1} = \sum_i \sum_j \frac{(x_i - x_j)^2/2}{v(v-1)} = \mathcal{E}(x_i - x_j)^2/2$$

et de la population des $x_1, x_2, \dots, x_1, \dots, x_v$, on en déduit la "super-population" des couples $x_i x_j (i \neq j)$.

L'ensemble des (i, j) étant décrit par le seul indice k

$$1 \leq k \leq \frac{v(v-1)}{2}$$

on posera :

$$y_k = (x_i - x_j)^2/2$$

Tout ceci amène à deux définitions :

(1) de la variance σ^2

$$v\sigma^2/v-1 = \bar{y} = \sum_k y_k / [v(v-1)/2]$$

(2) de l'estimation S^2 de $v\sigma^2/v-1$:

$$nS^2/n-1 = \bar{Y} = S_k y_k / [n(n-1)/2]$$

avec

$$\mathcal{E} \bar{Y} = \bar{y}$$

Variance de $\bar{Y}$.

a) Cas des tirages bernoulliens

Le calcul direct de $\mathcal{V} \bar{Y}$ (Réf. T3 par exemple) donne facilement :

$$\mathcal{V} \bar{Y} = \frac{\mu_4^0}{n} - \frac{(n-3)\sigma^4}{n(n-1)} = \frac{\mu_4^0 - 3\sigma^4}{n} + \frac{2\sigma^4}{n(n-1)}$$

où μ_4^0 est le moment des $(x_i - x)^4$.

$\mathcal{V} \bar{Y}$ est donc la somme de deux composantes, qu'on peut représenter par des divisions sur deux axes :

- l'une homothétique de $\mathcal{V} \bar{X}$ (1, 1/2, 1/3, ... 1/n ...)
- l'autre nouvelle (∞ , 1, 1/3, 1/6, 1/10, ... 2/n(n-1) ...)

b) Cas des tirages exhaustifs (Réf. T.C.R. 1, 1957).

Cette variance a été déjà calculée par de nombreux auteurs, dont Tukey (Jour. Am. Stat. Assoc. 1950, p. 501 & Ann. Math. Stat. 1956, p. 722), Chartier (Bull. Inf. I. N. S. E. E., Mars 1957, p. 31)...

Ces calculs se simplifient notablement si on a recours à la définition suivante de la variance

$$\sigma^2 = \sum (x_i - x_j)^2 / v^2$$

ou, pour abrégé
$$\sigma^2 = \sum (a-b)^2 / v^2 = \frac{v-1}{2v} \mathcal{E}(a-b)^2$$

et si, par ailleurs, on remarque que, de l'identité

$$(a-b)^2(a-c)^2 + (b-c)^2(b-a)^2 + (c-a)^2(c-b)^2 \equiv \frac{1}{2} [(a-b)^4 + (b-c)^4 + (c-a)^4]$$

il suit qu'on a :

$$\mathcal{E}[(a-b)^2(a-c)^2] = \frac{1}{2} \mathcal{E}(a-b)^4$$

Dénombrons alors les termes du développement de $(\sigma^2)^2 = \sigma^4$

$$\text{Termes en } \left\{ \begin{array}{lll} (a-b)^4 & ; & (a-b)^2(a-c)^2 & ; & (a-b)^2(c-d)^2 \\ M & + & 2M(v-2) & + & M \frac{(v-2)(v-3)}{2} \end{array} \right.$$

et, en remplaçant v par n et M par m , on passera du développement de $[\sum (a-b)^2]^2$ à celui de $[S(a-b)^2]^2 = (n^2 S^2)^2$. D'où les deux formules

$$v^4 \sigma^4 = \frac{v(v-1)^2}{2} \mathcal{E}(a-b)^4 + \frac{v(v-1)(v-2)(v-3)}{4} \mathcal{E}[(a-b)^2(c-d)^2]$$

$$\mathcal{E}[n^2 S^2]^2 = \frac{n(n-1)^2}{2} \mathcal{E}(a-b)^4 + \frac{n(n-1)(n-2)(n-3)}{4} \mathcal{E}[(a-b)^2(c-d)^2]$$

b. 1) En éliminant $\mathcal{E}[(a-b)^2(c-d)^2]$ entre ces formules et en écrivant :

$$\begin{aligned} \frac{1}{v^2} \sum_i \sum_j (x_i - x_j)^4 &= 2(\mu_4 - 4\mu_3\mu_1 + 3\mu_2^2) \\ &\equiv 2(\mu_4^0 + 3\sigma^4) \end{aligned}$$

d'où

$$\mathcal{E}(a-b)^4 \equiv \frac{v}{v-1} (\mu_4^0 + 3\sigma^4) \quad ,$$

en considérant par ailleurs les n valeurs z_i :

$$z_i = (x_i - \mu_1)^2$$

dont la moyenne $\bar{Z} = S(x_i - \mu_1)^2 / n$ estime $\zeta = \sigma^2$ avec une variance égale à :

$$\varphi \bar{Z} = \frac{\sigma_z^2}{n} \frac{v-n}{v-1}$$

avec

$$\sigma_z^2 = \mu_4^0 - \sigma^4$$

on retrouve l'expression habituelle de $\varphi \bar{Y}$

$$\varphi \bar{Y} = A \varphi \bar{Z} + B \sigma^4$$

avec

$$A = \frac{v(v-1)}{(v-2)(v-3)} \left[1 - \frac{2}{(n-1)(v-1)} \right]$$

avec

$$B = \frac{v(v-n)}{(v-1)(v-3)} \left[1 - \frac{n}{v-1} \right] \frac{2}{n(n-1)}$$

b. 2) Il est plus intéressant d'écrire :

$$\begin{aligned} \varphi \bar{Y} &= \varphi \left[\frac{nS^2}{n-1} \right] = \frac{\mathcal{E} n^4 S^4}{n^2(n-1)^2} - \frac{v^4 \sigma^4}{v^2(v-1)^2} \\ &= \frac{1}{2} \left(\frac{1}{n} - \frac{1}{v} \right) \mathcal{E} (a-b)^4 + \frac{1}{4} \left[\frac{(n-2)(n-3)}{n(n-1)} - \frac{(v-2)(v-3)}{v(v-1)} \right] \mathcal{E} [(a-b)^2(c-d)^2] \end{aligned}$$

$$\text{et comme } \frac{(n-2)(n-3)}{n(n-1)} = 1 - \frac{4}{n} + \frac{2}{n(n-1)}$$

$$\varphi \bar{Y} = \left[\frac{1}{2} \mathcal{E} (a-b)^4 - \mathcal{E} [(a-b)^2(c-d)^2] \right] \left(\frac{1}{n} - \frac{1}{v} \right) + \frac{1}{2} \mathcal{E} [(a-b)^2(c-d)^2] \left[\frac{1}{n(n-1)} - \frac{1}{v(v-1)} \right]$$

On retrouve deux composantes, avec (lorsque n varie) les mêmes divisions que dans le cas Bernoullien (qui correspond à v infini). On verra que c'est là un phénomène général.

Remarque.

$\varphi \bar{Y}$ n'est défini que si $n \geq 2$. On dira que c'est une perte d'information d'ordre 2.

9/ - Extension au cas de plusieurs variables.

a) Considérons X et X' deux variables quelconques (et leur distribution sur les unités de sondage (boules) avec :

$$\mathcal{E} x_i = \mu_1$$

$$\mathcal{E} x'_i = \mu'_1$$

$$\bar{X} = S x_i / n,$$

$$\bar{X}' = S x'_i / n$$

On a formellement :

$$\mathcal{V}(\bar{X} + t \bar{X}') = \frac{\alpha}{n} \cdot \frac{v - n}{v - 1}$$

en posant :

$$\begin{aligned} \alpha &= \mathcal{E} [x_i - \mu_1 + t(x'_i - \mu'_1)]^2 \\ &= \mathcal{E}(x_i - \mu_1)^2 + 2t \mathcal{E}[(x_i - \mu_1)(x'_i - \mu'_1)] + t^2 \mathcal{E}(x'_i - \mu'_1)^2 \end{aligned}$$

Ondira que $\mathcal{V}$ est une perte d'information correspondant à l'estimation simultanée de μ_1 et μ'_1 par l'échantillon d'effectif n .

b) Le terme rectangle de α possède (comme α lui-même et chacun des termes carrés) deux expressions, avec les écarts aux moyennes d'une part, avec les écarts $(x_i - x_j)$ et $(x'_i - x'_j)$ d'autre part :

$$2t \mathcal{E}(x_i - \mu_1)(x'_i - \mu'_1) = \alpha - \mathcal{E}(x_i - \mu_1)^2 - t^2 \mathcal{E}(x'_i - \mu'_1)^2$$

$$\frac{2t}{v^2} \sum_{i,j} (x_i - x_j)(x'_i - x'_j) = \alpha - \frac{1}{v^2} \sum \sum (x_i - x_j)^2 - \frac{t^2}{v^2} \sum \sum (x'_i - x'_j)^2$$

Il en résulte pour la covariance de x_i et x'_i des Lemmes analogues à ceux de la variance.

c) Comparons les deux formules (où les x_i et x'_i sont supposés essentiellement positifs),

$$\mathcal{E}(x_i - \mu_1)^2 = \mathcal{E}x_i^2 - \mu_1^2$$

$$\mathcal{E}(x_i - \mu_1)(x'_i - \mu'_1) = \mathcal{E}(x_i x'_i) - \mu_1 \mu'_1 = \mathcal{E}(\sqrt{x_i x'_i})^2 - (\sqrt{\mu_1 \mu'_1})^2$$

La covariance est donc une perte d'information pour l'estimation de $\sqrt{\mu_1 \mu'_1}$.

d) On peut employer $\mathcal{V}(Z + t Z')$ ou $\mathcal{V}(uZ + vZ' + wZ'')$ comme perte d'information dans des cas beaucoup plus généraux : estimateurs sans biais Z, Z', Z'' ... de caractéristiques ζ, ζ', ζ'' ... d'une distribution à une ou plusieurs variables.

Exemple.

Estimation simultanée des deux premiers moments :

$$\text{avec} \quad = \mathcal{E}[(x_i^2 - \mu_2) + t(x_i - \mu_1)]^2$$

$$\alpha = \mathcal{E}(x_i^2 - \mu_2)^2 + 2t\mathcal{E}(x_i^2 - \mu_2)(x_i - \mu_1) + t^2\mathcal{E}(x_i - \mu_1)^2$$

10/ - Urne avec plus de deux couleurs de boules.

Lorsqu'on estime les proportions (p q r) d'unités de sondage possédant les caractères $\mathcal{A}$ $\mathcal{B}$ $\mathcal{C}$ s'excluant mutuellement ($p + q + r = 1$), il serait utile de savoir calculer une "distance" entre les points (pqr) et (PQR) dans le plan d'équation

$$x + y + z = 1$$

Pour cela on formera :

$$V = \mathcal{V}(uP + vQ + wR)$$

$$= \mathcal{V}[(u - w)P + (v - w)Q]$$

$$= \frac{v - n}{n(v - 1)} [(u - w)^2 p(1 - p) + (v - w)^2 q(1 - q)] + 2(u - w)(v - w) [\mathcal{E}PQ - pq]$$

Mais on sait que :

$$\frac{v}{v - 1} \mathcal{E}PQ = \frac{n}{n - 1} pq$$

$$\text{D'où : } V = \frac{v}{v - 1} [p(1 - p) - 2tpq + t^2q(1 - q)] \left(\frac{1}{n} - \frac{1}{v} \right)$$

(avec $t = v - w / u - w$).

11/ - Extension à certaines estimations biaisées.

Moyennes généralisées.

A chaque unité i de la population, faisons correspondre y_i , $y_i = f(x_i)$, $f(x)$ désignant une fonction connue arbitraire.

Utilisons le concept de moyenne généralisée (GINI)⁽¹⁾; soit (x^*) la moyenne généralisée des x_i relativement à la fonction f , définie par l'équation :

$$f(x^*) = \bar{y} \quad \text{ou} \quad x^* = f^{-1}(\bar{y})$$

L'échantillon fournit une estimation (X^*) de (x^*) , définie par l'équation :

(1) Réf. : par exemple T2.

$$f(X^*) = \bar{Y} \quad \text{ou} \quad X^* = f^{-1}(\bar{Y})$$

$$\mathcal{V} \bar{Y} = \frac{v}{v-1} \sigma_y^2 \left(\frac{1}{n} - \frac{1}{v} \right)$$

est la perte d'information relative à l'estimation de $\bar{y}$ par $\bar{Y}$, autrement dit : de $\sum f(x_i)/v$ par $S f(x_i)/n$.

Trois questions se posent :

- Peut-on dire également que c'est la perte d'information relative à l'estimation de x^* par X^* ?

- Ne doit-on au contraire considérer que $\mathcal{V} X^*$ est la perte d'information relative à l'estimation de x^* par X^* ?

- N'existe-t-il donc pas finalement plusieurs façons distinctes de représenter ladite perte d'information ?

Elles tiennent au fait que X^* n'est pas un estimateur sans biais de x^* quand $\mathcal{E} \bar{Y} - \bar{y} = 0$, sauf si f est linéaire.

12/ - Cas des estimateurs biaisés en général.

Le biais étant $b = Z - \zeta$

$$\begin{aligned} \text{il vient :} \quad \mathcal{V} Z &= \mathcal{E} [Z - \zeta + \zeta - \mathcal{E} Z]^2 \\ &= \mathcal{E} (Z - \zeta)^2 + b^2 - 2b^2 \end{aligned}$$

Lorsque l'effectif varie, on voit facilement que ni $\mathcal{E} Z$, ni $\mathcal{E} (Z - \zeta)^2$ ne satisfont à la relation de récurrence des informations. $\mathcal{E} Z^2 - \zeta^2$ conviendrait manifestement, mais rien ne prouve que ce soit là une perte d'information imputable à l'estimateur Z , pour l'estimation de ζ (voir § II Ch. IV).

Si l'on convient de mesurer la perte d'information relative à X^* par $\mathcal{V} \bar{Y}$, on n'est pas en possession d'une expression analogue pour les autres estimations biaisées (souvent beaucoup plus usuelles).

Il est courant dans la pratique des sondages - lorsque v est de l'ordre du million et n de l'ordre du millier - que $\mathcal{V} Z$ et b^2 ne diffèrent pratiquement pas d'infiniment petits de la forme

$$k/n \text{ et } k'/n^2;$$

de sorte que b^2 est négligeable à côté de $\mathcal{V} Z$. En confondant x^* et $\mathcal{E} X^*$ ou ζ et $\mathcal{E} Z \dots$, $\mathcal{V} X^*$ ou $\mathcal{V} Z$ est approximativement perte d'information.

On pourrait se demander s'il n'y a pas lieu de n'attacher la notion de perte d'information qu'aux estimateurs sans biais, vu que les estimateurs "biaisés" ne sont justement employés que dans les cas où on juge b^2 négligeable à côté de $\mathcal{V}X^*$. On reviendra sur cette question au §II, Ch.IV.

13/ - Cas d'une population infiniment nombreuse.

Revenons sur le cas des sondages avec remise (§ 3°, 4°, 5°) et supposons v infiniment grand: rien ne s'en trouvera changé. En particulier: La variable x peut avoir une loi de distribution telle que σ^2 n'existe pas, $\mathcal{V}X$ étant infiniment grand. Alors la notion de perte d'information semble en défaut.

On peut voir une raison de ne pas accepter la présente théorie. On cherchera alors à substituer à $\mathcal{V}Z = E(Z - \zeta)^2$ une expression de la forme $EW(Z, \zeta)$ qui reste finie cette fois. Mais on verra au chapitre IV que pareille expression n'a aucune raison d'être une information au sens de Schützenberger; si une même forme de W doit rester valable pour toutes les lois de distribution, et si Z estime ζ sans biais, on verra que (sous certaines conditions de régularité pour W), $(Z - \zeta)^2$ est à un facteur près la seule forme possible pour W .

Ceci n'exclut pas l'emploi de pertes d'information n'ayant pas la forme EW (comme on va le voir au chapitre II), non plus d'ailleurs que le recours à des expressions EW , où la forme même de W dépendrait étroitement de la loi de distribution.

Il semble plus logique d'appeler formellement $\mathcal{V}Z$ perte d'information, alors même que σ^2 serait infini, toutes les pertes (d'ordre I) étant définies (à une similitude près) par la division $1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{v}$.

CHAPITRE II

LES PERTES D'INFORMATION DANS LE CAS DE L'URNE DE BERNOULLI ⁽¹⁾

I - DEFINITION GENERALE DE LA PERTE EN SONDAGE EXHAUSTIF -

Les variables $X, Y, \dots$ prenant les valeurs $(x_i, y_i, \dots)$ sur la "boule" i de l'urne (v) d'effectif v , considérons un échantillon de n boules (tirées au sort sans remise avec d'égales probabilités) et une expression $p(v, n)$ des x_i (de l'univers et de l'échantillon),

nulle si $n = v$

définie positive pour n entier $v > n \geq r$

strictement croissante avec $(v - n)$.

On dira de toute expression $p(v, n)$ qu'elle représente une perte d'information due au fait qu'on ne connaît de l'urne (v) que l'échantillon (n) , si elle est liée à $p(v, n')$ correspondant à l'échantillon (n') , $n > n'$, et à $P(n, n')$ c'est-à-dire à la perte aléatoire d'information sur l'urne aléatoire (n) avec l'échantillon (n') , par la relation de récurrence :

$$p(v, n') = p(v, n) + \mathcal{E} P(n, n')$$

Remarque.

Il s'en suit que la variance $\mathcal{V}X$ est une perte d'information à ce nouveau point de vue (plus large que celui du chapitre I puisqu'il ne comporte explicitement aucune estimation d'un ζ par un certain Z).

II - LES PERTES D'INFORMATION D'ORDRE 1 -

1/ - Partons de

$$\frac{v\sigma^2}{v-1} \left(\frac{1}{n'} - \frac{1}{v} \right) = \frac{v\sigma^2}{v-1} \left(\frac{1}{n} - \frac{1}{v} \right) + \mathcal{E} \left[\frac{nS^2}{n-1} \left(\frac{1}{n'} - \frac{1}{n} \right) \right]$$

(1) Réf. T. C. R. 2.

relation établie au chapitre I.

Il résulte de la définition non courante de σ^2 et de S^2

$$\sigma^2 = \sum (x_i - x_j)^2 / v^2, \quad S^2 = S(x_i - x_j)^2 / n^2$$

que la relation de récurrence se conserve si l'on y remplace $\frac{1}{2} (x_i - x_j)^2$ par toute fonction non négative symétrique $f(x_i, x_j)$ telle que $f(x_i, x_j)$ soit nul.

D'où, pour que $p(v, n)$ soit une perte d'information il suffit que :

(Théorème)

$$\begin{aligned} p(v, n) &= \frac{2v}{v-1} \sum \frac{f(x_i, x_j)}{v^2} \left(\frac{1}{n} - \frac{1}{v} \right) \\ &= \mathcal{E}f(x_i, x_j) \cdot \left(\frac{1}{n} - \frac{1}{v} \right) \\ &= \gamma \left(\frac{1}{n} - \frac{1}{v} \right), \quad \text{avec } \gamma = \mathcal{E}f(x_i, x_j) \end{aligned}$$

Remarque.

Dans le cas particulier de

$$f = [g(x_i) - g(x_j)]^2$$

cette nouvelle perte d'information n'est autre que $\mathcal{V}\bar{Y}$, avec

$$y = g(x)$$

mais une interprétation aussi élémentaire fait défaut dans le cas où par exemple

$$f = g[(x_i - x_j)^2]$$

2/ - Notion d'ordre d'une perte d'information.

Dans le cas de l'estimation de la moyenne $\bar{x}$ de l'urne par celle $\bar{X}$ de l'échantillon, un échantillon formé d'une seule boule apporte déjà des renseignements. Aussi dirons-nous que $\mathcal{V}\bar{X}$, ainsi que les expressions

$$p(v, n) = \gamma \left(\frac{1}{n} - \frac{1}{v} \right)$$

avec

$$\gamma = \mathcal{E}f(x_i, x_j)$$

sont des pertes d'information du 1er ordre.

Les variances de l'estimation sans biais de σ^2 ou du coefficient de corrélation, par exemple, sont des pertes d'information, mais ne sont pas d'ordre 1.

3/ - Représentation graphique des pertes d'ordre 1.

Si on porte, sur un axe $\mathcal{V}\bar{X}$ ou toute autre perte $p(v, n)$ ainsi définie, on obtient un "axe d'information" (Δ) où, pour une urne (v) donnée, les échantillons des diverses tailles sont figurés par une division. Ces divisions sont semblables entre elles, quelles que soient la variable X , la fonction f et la taille v de l'urne. A l'échelle près, il n'existe donc qu'un axe (Δ).

4/ - Extension de la notion de perte d'ordre 1.

THEOREME -

Si, par hypothèse, $p(v, 1)$ est de la forme $\theta(v) \mathcal{E}f(x_i, x_j)$, on a nécessairement :

$$p(v, n) = [\theta(v) - \theta(n)] \mathcal{E}f(x_i, x_j)$$

En effet à :

$$p(v, 1) = \theta(v) \cdot \mathcal{E}^{(\nu)} f$$

correspond :

$$P(v-1, 1) = \theta(v-1) \mathcal{E}^{(\nu-1)} f$$

etc

Passons de l'état : v boules connues à l'état : 1 boule connue. On peut le faire :

- soit d'un seul coup (perte d'information $p(v, 1)$);

- soit en deux temps en passant par l'état : $(v-1)$ boules connues, d'où une perte d'information $p(v, v-1)$ suivie de la perte de $(v-2)$ boules sur $(v-1)$, c'est-à-dire espérance mathématique de la perte d'information

$$P(v-1, 1)$$

relative à une urne d'effectif $(v-1)$ mais de composition aléatoire.

D'où :

$$p(v, 1) = p(v, v - 1) + \mathcal{E} P(v - 1, 1)$$

c'est-à-dire :

$$\begin{aligned} p(v, v - 1) &= p(v, 1) - \mathcal{E} P(v - 1, 1) \\ &= [\theta(v) - \theta(v - 1)] \gamma \end{aligned}$$

Pour passer du cas où $n = v - 1$ au cas général où n est quelconque, il suffit de donner à l'échantillon des effectifs décroissants : $(v - 1), (v - 2), \text{etc.} \dots (n + 2), (n + 1), n$.

Les pertes d'information s'ajoutent, et il vient :

$$p(v, n) = [\theta(v) - \theta(n)] \cdot \gamma$$

Remarque.

Pour $n = 1$, on doit avoir nécessairement $\theta(1) = 0$.

En particulier : avec $\theta(v) = \frac{v - 1}{v}$; on retrouve les pertes précédentes :

$$\theta(v) - \theta(n) = \frac{1}{n} - \frac{1}{v}$$

Toute fonction positive décroissante $\theta(v)$ fournit des pertes d'information, qui ne diffèrent pas essentiellement entre elles et seront dites encore d'ordre 1, à condition que $\theta(v)$ soit définie pour toute valeur entière de v , avec $\theta(1) = 0$.

On remarquera pourtant que :

$$\left[1 - \frac{1}{v} \right] \mathcal{E} f = \frac{2}{v^2} \sum_{i > j} f(x_i, x_j) = \frac{1}{v^2} \sum_i \sum_j f(x_i, x_j) = \mathcal{E} 'f$$

$\mathcal{E}'$ désignant l'espérance mathématique étendue à toutes les combinaisons 2 à 2 avec répétition des $x_i x_j$, c'est-à-dire aux échantillons bernoulliens de deux boules.

On verra (§ V) que c'est la nécessité d'un raccord avec les sondages bernoulliens qui détermine les fonctions θ et, pour $r = 1$, conduit à prendre $\theta = 1 - \frac{1}{v}$.

III - LES PERTES D'INFORMATION D'ORDRE 2 -

1/ - Exemple de perte du 2ème ordre.

Avec un échantillon de taille $n \geq 2$, on peut estimer $\zeta = v\sigma^2/v - 1$, par $Z = nS^2/n - 1$; on sait exprimer $\mathcal{V}Z$ comme combinaison linéaire de $\mathcal{E}(x_i - x_j)^4$ et $\mathcal{E}[(x_i - x_j)^2(x_k - x_l)^2]$; à savoir :

$$\mathcal{E} \frac{(n^2 S^2)^2}{n^2(n-1)^2} = \frac{1}{n} \mathcal{E} \frac{(a-b)^4}{2} + \frac{(n-2)(n-3)}{n(n-1)} \mathcal{E} \frac{(a-b)^2 \cdot (c-d)^2}{4}$$

$$\left(\frac{v\sigma^2}{v-1} \right)^2 = \frac{1}{v} \mathcal{E} \frac{(a-b)^4}{2} + \frac{(v-2)(v-3)}{v(v-1)} \mathcal{E} \frac{(a-b)^2 \cdot (c-d)^2}{4}$$

$$\begin{aligned} \mathcal{V}Z &= \left(\frac{1}{n} - \frac{1}{v} \right) \mathcal{E} \frac{(a-b)^4}{2} - \left[\frac{2n-3}{n(n-1)} - \frac{2v-3}{v(v-1)} \right] \mathcal{E} \frac{(a-b)^2(c-d)^2}{2} \\ &= \left(\frac{1}{n} - \frac{1}{v} \right) \left[\frac{\mathcal{E}(a-b)^4}{2} - \mathcal{E}(a-b)^2(c-d)^2 \right] + \left[\frac{1}{n(n-1)} - \frac{1}{v(v-1)} \right] \mathcal{E} \frac{(a-b)^2(c-d)^2}{2} \end{aligned}$$

$\mathcal{V}Z$ est une perte d'information qui sera dite du 2ème ordre. C'est la somme d'une perte du 1er ordre et d'une du 2ème ordre : (car ni $(2n-3)/n(n-1)$, ni $1/n(n-1)$ ne sont définis pour $n < 2$).

2/ - Obtention de pertes plus générales.

Si l'on substitue $f(x_i, x_j)$ à $(x_i - x_j)^2$ avec : $f(x_j, x_i) = f(x_i, x_j) > 0$, $f(x_i, x_i) = 0$, le transformé de $\mathcal{V}Z$ est toujours une perte d'information d'ordre 2, puisque de la forme $A \left(\frac{1}{n} - \frac{1}{v} \right) + B \left(\frac{1}{n(n-1)} - \frac{1}{v(v-1)} \right)$.

3/ - Représentation graphique.

On considérera deux vecteurs colinéaires, représentant le terme en $1/n$ et le terme en $1/n(n-1)$, et par conséquent deux axes $\Delta 1$ et $\Delta 2$, portant respectivement les divisions en $1/n$ et $1/n(n-1)$, leurs échelles dépendant de la variable X et leurs origines de la taille v de l'urne.

4/ - Les cas où $n = 3$ et $n = 2$.

Allant de soi avec $n \geq 4$, le calcul de $\mathcal{V}Z$ serait en défaut pour $n = 3$ ou 2, si l'on ne raisonnait par continuité.

Pour $n = 3$, il vient :

$$\lim \mathcal{E} (a-b)^2(c-d)^2 \equiv \frac{2}{3} [(a-b)^2(a-c)^2 + (b-c)^2(b-a)^2 + (c-a)^2(c-b)^2]$$

(en faisant tendre d vers a , b ou c avec des probabilités $1/3$).

Etablissons l'identité :

$$(a-b)^2(a-c)^2 + (b-c)^2(b-a)^2 + (c-a)^2(c-b)^2 \equiv \frac{1}{2} [(a-b)^4 + (b-c)^4 + (c-a)^4]$$

Démonstration.

Posons $\alpha = a - b$; $\beta = b - c$; $\gamma = c - a$ et développons $(\alpha + \beta + \gamma)^4 \equiv 0$
Il vient :

$$\begin{aligned} & \alpha^4 + \beta^4 + \gamma^4 + (4 + 2)(\alpha^2\beta^2 + \beta^2\gamma^2 + \gamma^2\alpha^2) \\ & + (8 + 4)(\alpha^2\beta\gamma + \beta^2\gamma\alpha + \gamma^2\alpha\beta) + 4[\alpha^3(\beta + \gamma) + \beta^3(\gamma + \alpha) + \gamma^3(\alpha + \beta)] \equiv 0 \end{aligned}$$

Mais

$$\begin{cases} \beta + \gamma & = -\alpha & \text{etc} \\ \alpha^2\beta\gamma + \dots & = \alpha\beta\gamma (\alpha + \beta + \gamma) = 0 \end{cases}$$

Il reste ainsi :

$$(1 - 4)(\alpha^4 + \beta^4 + \gamma^4) + 6(\alpha^2\beta^2 + \beta^2\gamma^2 + \gamma^2\alpha^2) \equiv 0$$

D'où $\alpha^2\beta^2 + \beta^2\gamma^2 + \gamma^2\alpha^2 \equiv \frac{1}{2}(\alpha^4 + \beta^4 + \gamma^4)$

d'où $\lim \mathcal{E}(a - b)^2(c - d)^2 \equiv \mathcal{E}(a - b)^4$

Pour $n = 2$: cette dernière identité conserve un sens.

5/ - Conclusion.

$\forall Z$ est proportionnelle à $\mathcal{E}(a - b)^4$ pour $n = 3$ et $n = 2$, ce que le calcul direct montre (bien entendu) en faisant le changement de population :

$$\begin{aligned} y_k &= (x_i - x_j)^2/2 & 1 \leq i \leq v \\ & & 1 \leq k \leq \frac{v(v-1)}{2} \end{aligned}$$

On estime $v\sigma^2/v - 1$ par la moyenne $\bar{Y}$ des y échantillon, ce qui permet de trouver pour $n = 2$

$$\left(\frac{v-1}{v}\right)^v \mathcal{E}(\bar{Y}^2) = \frac{v-1}{2v} \frac{2 \sum (x_i - x_j)^4}{v^2}$$

La variance s'en déduit en retranchant σ^4 .

Pour $n = 3$, on trouve de même facilement :

$$\left(\frac{v-1}{v}\right)^2 \mathcal{E}(\bar{Y}^2) = \frac{v-1}{3v} \frac{2 \sum (x_i - \bar{x})^4}{v^2}$$

Pour les pertes plus générales, il n'en est pas ainsi; on a seulement :

$$\lim \mathcal{E}[f(a, b) f(c, d)] \equiv 2 \mathcal{E}[f(a, b) f(a, c)]$$

les composantes en $f(a, b)^2$ et $f(a, b) f(a, c)$ ne fusionnant pas; car on ne peut transposer l'identité :

$$(a-b)^2(a-c)^2 + (b-c)^2(b-a)^2 + (c-a)^2(c-b)^2 \equiv 2 [(a-b)^4 + (b-c)^4 + (c-a)^4]$$

Mais néanmoins les pertes sont définies jusqu'à $n = 2$ inclus.

IV - LES PERTES D'INFORMATION D'ORDRE QUELCONQUE r -

1/ - THEOREME 1 -

Si $p(v, n)$ est défini pour $n \geq r$, et si $p(v, r)$ est de la forme $\theta(v) \cdot \gamma$ où $\theta(v)$ croît avec v à partir de $\theta(r) = 0$, on a nécessairement :

$$p(v, n) = [\theta(v) - \theta(n)] \gamma,$$

lorsque l'expression C , isomorphe de γ pour l'échantillon (γ étant définie sur l'univers), satisfait à la relation :

$$\mathcal{E} C = \gamma$$

Démonstration.

Refaire le calcul du § I ci-dessus.

Exemple.

$\gamma = \mathcal{E}[(a-b)^2 \cdot (c-d)^2]$ avec $n \geq 4$ et passage à la limite pour $n=3$ et 2.

2/ - Il est évident que :

THEOREME 2 -

- Si p est une perte d'information d'ordre r , il en est de même de λp , λ étant une constante.

- Si p et p' sont des pertes d'information d'ordres respectifs r et r' , $\underline{r > r'}$, $(p + q)$ est une perte d'information d'ordre r .

- Toute combinaison linéaire de pertes d'information est

une perte d'information; et son ordre est le plus grand des ordres des composantes. Toutefois, si certains coefficients sont négatifs, il est nécessaire que :

- ladite combinaison linéaire ne devienne pas négative;
- les termes de rang le plus élevé ne se détruisent pas.

Réciproque du théorème 2.

On verra (§ III du Ch. IV) que, pour des plans de sondage très généraux (dont fait partie l'urne de Bernoulli) les pertes d'information sont nécessairement de la forme

$$\sum_i \sum_j [\theta_{ij}(v_i) - \theta_{ij}(n_i)] \gamma_{ij}$$

3/ - Sur quelques difficultés relatives à l'ordre d'une perte d'information.

a) Il est clair que l'ordre est déterminé par la nature de la fonction $\theta(v)$; mais l'expression elle-même n'est définie qu'autant que n et v ne descendent pas en-dessous d'une limite s qui ne coïncide pas nécessairement avec r .

Ainsi $(a-b)^2(c-d)^2$ correspondait à $s=4$. Par continuité, on est passé de là à $n=3$, puis $n=2$; étant entendu que $\theta(n)$ serait infini si l'on allait jusqu'à $n=1$ (donc $r=2$).

b) $\gamma = \mathcal{E}[(a-b)^2(a-c)^2 + (b-c)^2(b-a)^2 + (c-a)^2(c-b)^2]$ est en réalité (on l'a vu) : $\gamma = \frac{1}{2} \mathcal{E}(a-b)^4$; de même

$$\mathcal{E}[(a-b)(a-c) + (b-c)(b-a) + (c-a)(c-b)]$$

n'est autre que :

$$\gamma = \frac{1}{2} \mathcal{E}(a-b)^2,$$

car on a :

$$(a-b)(a-c) + (b-c)(b-a) + (c-a)(c-b) = \frac{1}{2} [(a-b)^2 + (b-c)^2 + (c-a)^2]$$

Considérons, au contraire, l'expression :

$$|a-b| \cdot |b-c| + |b-c| \cdot |c-a| + |c-a| \cdot |a-b|$$

Si l'on suppose $a > b > c$, elle n'est autre que :

$$(a - c)^2 + (a - b)(b - c)$$

Ainsi la valeur de s n'est pas toujours évidente; tandis qu'aucun doute n'existe pour r , une fois la fonction θ choisie.

c) Mode d'obtention des expressions C et γ ci-dessus.

Considérons p valeurs de x (4, par exemple) obtenues par tirage exhaustif, désignées par $x_1, x_2, \dots, x_1$.

Désignons par $f(i, j \dots 1)$ une fonction positive symétrique des $x_1, x_2, \dots, x_1$;

Désignons par $\mathcal{E}_0$ l'espérance mathématique étendue aux combinaisons p à p des v boules;

Désignons par E l'espérance mathématique étendue aux combinaisons p à p de (n) boules, elles-mêmes tirées exhaustivement dans l'urne.

Posons : $\gamma = \mathcal{E}_0 f(i, j \dots 1)$ et $C = E f(i, j \dots 1)$. Alors il est bien clair qu'on a :

$$\mathcal{E}(C) = \gamma$$

du moins tant que $n \geq p$ (et pour $r \leq n < p$, on obtient par continuité les limites des pertes d'information).

d) Remarque.

On était naturellement conduit à définir, outre les pertes d'information du 1er ordre, des pertes d'information :

- du 2ème ordre, à base de fonctions $f(i, j, k)$
- du 3ème ordre, à base de fonctions $f(i, j, k, l)$,
- etc.

Tenons-nous-en au 2ème ordre. Nous avons commencé comme suit :

On définissait $f(i, j, k)$ - perte d'information lorsque deux des trois boules sont connues - comme une fonction positive symétrique de x_1, x_2, x_k :

$$f(i, j, k) = f(i, k, j) = \dots \text{ etc.}$$

Il y a une chance sur trois pour que x_1 ne soit pas tiré, x_2 et x_k

ayant été tirés; la perte d'information correspondante serait notée :

$$(i; jk) = (i; kj)$$

On aurait donc postulé que :

$$f(i j k) = \frac{1}{3} [(i; jk) + (j; ki) + (k; ij)]$$

Après quoi il était fort tentant de poser, pour un univers d'effectif quelconque :

$$\gamma = \mathcal{E}f(i j k)$$

On aurait donc eu :

$$\begin{aligned} & - \text{ si } x \text{ prend les valeurs } a, b, c, \text{ (pour } v = 3); \\ & \qquad \qquad \qquad a, b, c, d, \text{ (pour } v = 4); \end{aligned}$$

$$\gamma = f(abc) \qquad \qquad \qquad \text{dans le 1er cas;}$$

$$\gamma = \frac{f(abc) + f(abd) + f(acd) + f(bcd)}{4} \quad \text{dans le 2ème cas.}$$

Mais nous avons rencontré l'expression

$$(a - b)^2(c - d)^2 + (a - c)^2(b - d)^2 + (a - d)^2(b - c)^2$$

qui est une certaine perte d'information correspondant au cas où $v = 4$, $n = 2$ et qu'on ne peut mettre sous la forme :

$$f(abc) + f(abd) + f(acd) + f(bcd)$$

La base de départ était donc trop étroite.

e) Image géométrique.

De même que, pour l'ordre 1, $f(x_i, x_j)$ représente une distance entre x_i et x_j , il est naturel de voir dans $f(ijk)$ le demi-périmètre du triangle (ijk) , qui se réduit à la distance $f(ik)$ si i et j viennent se confondre. Ce serait une erreur de voir dans $f(ijk)$ l'aire d'un triangle, dans $f(ijkl)$ le volume d'un tétraèdre ; car cela conduirait à postuler :

$$f(i i k) = 0 \quad \text{etc.}$$

L'image du périmètre incite pourtant à écrire :

$$f(ijk) = \frac{1}{2} [(ij) + (jk) + (ki)]$$

avec
(ij) = f(ijk), etc.

$$= \frac{3}{2} \cdot \frac{(ij) + (jk) + (ki)}{3}$$

expressions de pertes du 1er ordre, fallacieusement présentées sous un aspect du 2ème ordre, ce qui ferait croire que tout f(ijk) peut se mettre sous cette forme.

V - TIRAGES AVEC REMISE DANS L'URNE -

1/ - La définition de la perte, posée dans le cas des tirages exhaustifs, se modifie avec les tirages bernoulliens comme suit; cette expression $p'(v, n)$:

- est définie positive pour n entier, $n \geq r$;
- est strictement décroissante quand n croît;
- a une limite nulle pour n infini;
- satisfait à la relation de récurrence suivante :

$$p'(v, n) - p'(v, n+1) = \mathcal{E}'P(n+1, n)$$

où $\mathcal{E}'$ désigne l'espérance en tirage bernoullien, et P la perte d'information dans le cas où l'on connaît n boules sur (n+1) d'une urne (de composition aléatoire); P est calculée aux paragraphes ci-dessus.

En substituant à $P(n+1, n)$ l'expression

$$[\theta(n+1) - \theta(n)]C_n$$

avec $\mathcal{E}C = \gamma$, $\theta(r) = 0$, $\theta(n) < \theta(n+1)$, il vient

$$p'(v, n) = \sum_n^{+\infty} [\theta(h+1) - \theta(h)] \mathcal{E}'C_h$$

2/ - Calcul de $\mathcal{E}'C_h$.

a) Pertes d'ordre 1.

$$C_n = Ef(x_i, x_j)$$

$$\mathcal{E}'C_n = \mathcal{E}'Ef = E\mathcal{E}'f$$

$$= \mathcal{E}'f = \frac{v-1}{v} \gamma$$

d'où :

$$p'(v, n) = [\theta(\infty) - \theta(n)] \frac{v-1}{v} \gamma$$

En particulier
$$p'(v, 1) = \theta(\infty) \frac{v-1}{v} \gamma$$

Il faut se souvenir en outre que le tirage bernoullien est indiscernable du tirage exhaustif pour $n = 1$:

$$p'(v, 1) = p(v, 1)$$

d'où $\frac{v}{v-1} \theta(\infty) = \theta(v)$, c'est-à-dire [si l'on pose $\theta(\infty) = 1$]

$$\theta(v) = 1 - \frac{1}{v}$$

b) Cas général.

THEOREME -

Si les espérances mathématiques d'une statistique donnée Y, pour des tirages avec et sans remise dans l'urne, $\mathcal{E}'Y$ et $\mathcal{E}Y$ ont des limites, quand l'effectif de l'urne tend vers l'infini, ces limites coïncident en général.

En effet : En n tirages équiprobables avec remise, la probabilité de tirer n boules distinctes est $\left(1 - \frac{1}{v}\right) \left(1 - \frac{2}{v}\right) \dots \left(1 - \frac{n-1}{v}\right)$ qui tend vers 1 si v tend vers l'infini, n restant constant. La probabilité des valeurs non nulles de $(\mathcal{E}'Y - \mathcal{E}Y)$ tend vers zéro. Si l'on suppose que Y reste fini on voit que, si $\mathcal{E}Y$ tend vers une limite, $\mathcal{E}'Y$ tendra vers la même; et vice versa.

Corollaire 1.

Partant d'une infinité d'urnes(v) de même composition, et mélangeant 1, 2, 3, ... λ ... urnes identiques, tirant chaque fois n boules avec remise d'une part, sans remise de l'autre, - calculant $\mathcal{E}'C_n$ et $\mathcal{E}C_n$, avec

$$C_n = Ef(i \ j \ \dots \ 1);$$

alors, si $\alpha = \mathcal{E}'f$ est une constante en λ , on a nécessairement

$$\lim_{v \rightarrow \infty} \mathcal{E}C_n = \alpha \quad (\text{sous réserve que } C_n < \infty)$$

Pour l'établir on notera que $\mathcal{E}'C_n = \mathcal{E}'(Ef) = E(\mathcal{E}'f)$

$$= \mathcal{E}'f = \alpha$$

Corollaire 2.

Dans les mêmes conditions, si $\mathcal{E}C_n$ a une limite α lorsque λ tend vers l'infini et si f ne dépend pas de v , on a nécessairement :

$$\mathcal{E}'f = \lim_{\nu \rightarrow \infty} \mathcal{E}C_n \quad (\text{sous réserve que } C_n < \infty)$$

Conséquence.

$$\text{De} \quad p(v, n) = \sum_{h=0}^{v-1} [\theta(h+1) - \theta(h)] \mathcal{E}C_h = [\theta(v) - \theta(n)] \gamma$$

$$\text{on déduit} \quad \sum_n [\theta(h+1) - \theta(h)] \lim_{\lambda \rightarrow \infty} \mathcal{E}C_h = [\theta(\infty) - \theta(n)] \lim_{\lambda \rightarrow \infty} \gamma$$

$$\text{D'où} \quad p'(v, n) = \lim_{\lambda \rightarrow \infty} p(\lambda v, n) = [\theta(\infty) - \theta(n)] \alpha \quad \text{avec} \quad \lim_{\lambda \rightarrow \infty} \gamma = \alpha$$

Exemple.

$3C = f(a, b) f(c, d) + f(a, c) f(b, d) + f(a, d) f(b, c)$ (en supposant $f < \infty$, f ne dépend pas de v).

$$\alpha = \mathcal{E}'C = \mathcal{E}'[f(a, b) f(c, d)]$$

$$= \left(\frac{v-1}{v}\right) \left(\frac{v-2}{v}\right) \left(\frac{v-3}{v}\right) \mathcal{E}[f(a, b) f(c, d)] + \bar{\omega} \mathcal{E}[f(a, b) f(a, c)] \\ + \dots$$

Remplaçons v par λv et faisons tendre λ vers l'infini. C ne dépend pas de l'effectif de l'urne; de même $\mathcal{E}'C$.

Il reste :

$$\alpha = \mathcal{E}'C = \lim_{\lambda \rightarrow \infty} \mathcal{E}[f(a, b) f(c, d)]$$

3/ - Relation entre tirages bernoulliens et tirages exhaustifs.

A effectif égal n pour l'échantillon, la perte p' est nécessairement supérieure à p .

$$p'(v, n) - p(v, n) > 0,$$

$$\text{ou} \quad [\theta(\infty) - \theta(n)] \alpha - [\theta(v) - \theta(n)] \gamma > 0$$

Pour $n = r$, il n'y a pas lieu d'identifier p et p' ; car il n'est pas

exact que la perte d'information soit là même avec r unités suivant qu'on les sait distinctes ou non.

Il en résulte apparemment qu'on a plus de latitude dans le choix de la fonction θ que pour l'ordre 1.

En pratique, on a vu (au § III) que, pour $r = 2$, avec

$$\gamma = E(a - b)^2(c - d)^2$$

l'expression suivante était la seule qui subsistât :

$$\theta(v) - \theta(n) = \frac{1}{n(n-1)} - \frac{1}{v(v-1)}$$

ce qui donne, pour obtenir une fonction $\theta(v)$ croissante, s'annulant avec $v = 2$,

$$\theta_2(v) = \frac{1}{1.2} - \frac{1}{v(v-1)} = \frac{v(v-1) - 2}{2v(v-1)}$$

On est arrivé à pareille expression par des calculs de nombres de termes. Plus généralement, au moins chaque fois qu'on se limitera à des tirages équiprobables, le calcul de θ fera intervenir l'analyse combinatoire, et sera une fraction rationnelle à coefficients entiers de variable v .

Une fraction rationnelle, croissante ($v \geq r$), admettant le zéro ($v = r$) et les pôles $v = r - 1, r - 2, \dots$ O, et dont les 2 termes sont de degré minimum (r); est

$$\theta_r(v) = \frac{1}{r!} - \frac{1}{v(v-1)\dots(v-r+1)} = \frac{1}{r!} - \frac{(v-r)!}{v!}$$

Une fraction rationnelle dont les 2 termes ne sont pas de même degré ne convient pas (du moment qu'on postule l'existence de $\theta(\infty)$).

Toute fraction rationnelle admettant ces pôles peut être décomposée en éléments simples du type précédent (plus un polynôme), soit :

$$a_0 + a_1 \theta_1(v) + a_2 \theta_2(v) \dots + a_r \theta_r(v)$$

(on le voit en identifiant). Le polynôme disparaît en supposant le numérateur de degré r . La fraction admet ($v = r$) pour zéro si les a_i sont liés par

$$0 = a_0 + a_1 \left(1 - \frac{1}{r}\right) + a_2 \left(\frac{1}{2!} - \frac{1}{r(r-1)}\right) + \dots + a_{r-1} \left(\frac{1}{r-1!} - \frac{1}{r!}\right)$$

relation qui détermine a_0 . Finalement la forme générale de $\theta(v)$ pour l'ordre r est cette combinaison $\sum a_i \theta_i(v) + a_0$ dans les hypothèses du présent chapitre.

Une théorie n'envisageant de tirages que dans une population infiniment nombreuses (comme celle de Lindley, voir chapitre VIII) n'exclurait pas, en outre, la présence d'un polynôme, rendant infini $\theta(\infty)$.

4/ - Cas d'une population infiniment nombreuse.

Il n'est pas exclu que $\lim \gamma(v \rightarrow \infty)$ n'existe plus pour certaines distributions (voir n° 13, Ch. I), suivant qu'on a choisi telle ou telle fonction $f(i, j, \dots)$; car des expressions uniformément bornées $|f(i, j)| < M$ n'auraient guère d'intérêt pratique.

CHAPITRE III

LA PERTE D'INFORMATION AVEC LES PLANS DE SONDAGE QUELCONQUES

I - DEFINITIONS FONDAMENTALES -

A - DEFINITION D'UN PLAN DE SONDAGE.

1/ - Structure de l'univers sondé.

Le problème essentiel des sondages est l'évaluation de certaines caractéristiques d'une population donnée, composée d'éléments.

Par exemple, pour une population humaine discontinue, l'élément est l'être humain; on rencontre aussi des populations "continues", tel le territoire agricole, pour qui on peut envisager l'are, le mètre carré, le centimètre carré, ... comme élément constitutif.

Ces caractéristiques sont (le plus souvent) des grandeurs additives, des rapports de grandeurs additives, etc. ce qui autorise les remaniements suivants.

A cette population en quelque sorte brute, la théorie des sondages substitue une population dotée d'une certaine structure. On sera déjà dans un cas très général en envisageant la structure suivante :

Un premier découpage A, en segments portant les numéros : 1, 2, ...
i ... a;

Un second découpage B, à l'intérieur de chacun des segments précédents, d'où constitution de sous-segments numérotés : i1, i2, ... ij ... ib; à l'intérieur du segment n°i;

puis un troisième, un quatrième découpage, etc. pour aboutir en dernier lieu aux "éléments" de la population portant un numéro : i, j, k, ... l.

avec

$$1 \leq i \leq a$$

$$1 \leq j \leq b_i$$

$$1 \leq k \leq c_{ij}$$

.....

$$1 \leq l \leq d_{ijk}$$

En pratique la structure est en partie imposée, en partie le résultat du choix plus ou moins judicieux du statisticien.

2/ - Cadre et structure de l'échantillonnage.

Une fois fixée la structure de la population à sonder, on donne le cadre de l'échantillonnage; dans le cas très général décrit ci-dessus, cela signifiera qu'on détermine le nombre des segments, sous-segments, ... éléments à conserver à l'échantillon, dans chacun des découpages.

Pratiquement on se donne en général :

- soit les effectifs $n_A, n_B \dots$ des unités de sondages à conserver;

- soit les fractions de sondage $f_A, f_B \dots$ proportions d'unités de sondage à conserver.

Les problèmes réels nécessitent qu'on distingue une notion intermédiaire entre la structure de la population et le cadre de l'échantillon; on conviendra de dire que : Fixer la structure de l'échantillon signifie décider quelles fractions de sondage sont égales à 1, les autres fractions restant à déterminer (entre 0 et 1, bornes exclues).

Pour abrégé, on parlera des effectifs ou de la taille d'un échantillon, pour désigner les paramètres n ou f restant à préciser lorsque la structure de l'échantillon est fixée.

3/ - Terminologie courante.

a) Si la fraction de sondage f_A retenue pour le premier découpage est égale à l'unité, la fraction f_B étant inférieure à 1, le découpage A est appelé stratification et chaque segment est appelé strate.

Chaque parcelle résultant du découpage B est appelée unité primaire de sondage (U. P. S. en abrégé) ou unité du 1er ordre.

Si, au contraire, la fraction f_A est inférieure à l'unité, c'est le découpage A qui procure les unités primaires de sondage.

b) Si (par exemple) f_b étant inférieur à 1, f_c est égal à 1, le découpage C est appelé sous-stratification; et chaque parcelle C est une sous-strate.

Chaque parcelle résultant du découpage D est appelée unité de sondage du 2ème degré.

Si, au contraire, la fraction f_c est inférieure à 1, le découpage C donne naissance aux unités du 2ème degré.

On définit de même les unités de sondage d'ordre 3, voire d'ordre quelconque.

Cette terminologie peu rationnelle s'explique par le fait que, dans la majorité des sondages réels, on a :

$$f_A = 1, \quad f_B < 1$$

4/ - Désignation de l'échantillon.

Le cadre de l'échantillonnage étant fixé, - pour définir le plan d'échantillonnage il reste à se donner le mode de désignation de l'échantillon.

On ne s'intéressera ici qu'aux échantillons tirés au sort⁽¹⁾, c'est-à-dire constitués comme suit :

- unités de sondage du 1er ordre tirées au sort (à l'intérieur des strates);

- unités de sondage du 2ème ordre tirées au sort parmi celles qui constituent les unités du 1er ordre tirées (et à l'intérieur des sous-strates);

- unités de sondage du 3ème ordre tirées au sort, etc.

Il convient de préciser le mode de tirage au sort et ceci pour chaque degré de sondage. Les modes les plus courants sont :

- sondage bernoullien (à la manière des boules de l'urne de Bernoulli) ou avec remise;

- sondage exhaustif (sans remettre les boules tirées dans l'urne);

(1) On évite de parler d'un échantillon aléatoire, en raison du sens péjoratif de ce mot en français courant.

- sondage bernoullien ou exhaustif avec probabilités inégales de tirage pour les diverses unités;

- sondage équilibré ("balanced sampling"); plus généralement le tirage au sort peut être assorti de clauses restrictives, destinées à écarter l'apparition des cas extrêmes.

5/ - Estimateurs et erreurs d'estimation.

Soit $x_{ijk...1}$ la valeur que prend une variable X sur l'élément $(ijk \dots 1)$.

Soit ζ tout paramètre à estimer sur un échantillon d'éléments $(ijk \dots 1)$. L'estimateur Z de ζ est une certaine fonction des x échantillons.

6/ - Coût d'une enquête sur échantillon.

Le coût des enquêtes est un facteur essentiel dans la pratique. Aussi introduit-on dans la théorie des sondages une fonction de coût⁽¹⁾ (plus ou moins réaliste), où entrent les divers effectifs de l'échantillon (à côté d'autres paramètres).

B - BUT DE LA THEORIE DES SONDAGES.

La théorie des sondages vise à permettre le choix du "meilleur" plan de sondage et du meilleur estimateur, sous des conditions restrictives telles que : pour un coût total donné, - ou pour certaines composantes du coût déterminées.

1/ - A la recherche du concept de "meilleur plan de sondage".

Toutes choses égales d'ailleurs, le meilleur plan de sondage serait celui qui fournit en moyenne les erreurs d'échantillonnage les plus faibles. A chaque ζ correspond, pour chaque Z , un écart $(Z - \zeta)$ erreur d'échantillonnage (ou d'estimation); on cherche à se faire une idée de l'ordre de grandeur de $(Z - \zeta)$ en définissant (et évaluant numériquement) des indicateurs d'erreur (ζZ) .

a) $Z - \zeta$ est le "biais" (bias) et représente l'erreur systématique d'échantillonnage.

Un estimateur présentant un biais important et inconnu (de façon à exclure toute correction de biais) serait inutile.

b) Avec un biais nul ou très faible, on est assuré que l'estimation sera "bien centrée"; mais il peut se faire qu'elle soit très

(1) Voir par exemple T. 7.

imprécise si des écarts $Z - \xi$ très importants ont des chances non négligeables de se produire. C'est pourquoi on définit des indicateurs de précision.

Le plus utilisé est la variance $VZ = \mathcal{E}(Z - \xi)^2$ ou $\mathcal{E}(Z - \xi)^2$ lorsque Z est sans biais.

Concurremment avec la variance on emploie encore l'erreur-type, le coefficient de variation, l'erreur absolue moyenne (voir ci-dessus, § I du Chap. I).

Au moins depuis Neyman (1934), les plans de sondage optimum de la théorie classique sont définis de façon à rendre minimum certaines variances d'échantillonnage. (minimum lié).

Nous dirons plus généralement que le meilleur plan de sondage est celui qui fait perdre le moins d'information (toutes choses égales d'ailleurs).

2/ - Caractères généraux de la perte d'information.

On convient d'exiger a priori d'une perte d'information les caractères suivants :

1) être nulle si l'échantillon coïncide avec la population;

2) être infinie si l'échantillon a un effectif nul;

3) être déterminée par le plan d'échantillonnage et le choix de l'estimateur;

4) être croissante lorsque, pour une structure donnée, la taille de l'échantillon décroît.

La théorie des sondages définit déjà certains indicateurs d'erreur attachés à un estimateur donné.

Il est naturel de se demander si ces indicateurs (ξZ) ne mesurent pas également la perte d'information.

Le biais, la variance, l'écart-type, le coefficient de variation, l'écart moyen présentent les quatre caractères ci-dessus.

Un cinquième caractère va décider si ce sont ou non des pertes d'information.

Donnons-nous la structure de l'échantillon, le mode de tirage et

l'estimateur. On a une famille (F) de plans de sondage ne différant entre eux que par la taille. Considérons deux plans de sondage (1) et (2) tels que la taille de (2) soit inférieure à celle de (1); ce qui signifie que l'un au moins des effectifs de (2) est inférieur à celui de (1) correspondant et qu'aucun n'est supérieur. Ainsi l'indicateur d'erreur (ζZ_2) est plus grand que l'indicateur (ζZ_1) , mais la différence $(\zeta Z_2) - (\zeta Z_1)$ n'a pas elle-même, en général, de signification intrinsèque.

Supposons qu'on ait, en outre, la relation de Chasles, plus précisément :

$$(\zeta Z_2) = (\zeta Z_1) + \mathcal{E}_1(Z_1 Z_2)$$

$\mathcal{E}_1$ désignant l'espérance mathématique étendue à toutes les valeurs de l'aléatoire Z_1 substitué à ζ . Alors il est naturel de dire : (ζZ) est perte d'information relative à l'estimation de ζ (axiomatique de Schutzemberger).

Toutefois si, dans le cas de tirages dans une urne, avec $\mathcal{E}Z = \zeta$ $\forall Z = (\zeta Z)$ possède les cinq caractères ci-dessus, cela cesse d'être vrai pour le plus banal des sondages à deux degrés.

Cas d'un sondage à deux degrés et deux paramètres.

Les unités étant tirées avec d'égales probabilités et exhaustivement, à chacun des degrés de sondage, - les tirages au second degré sont supposés indépendants du résultat au premier degré.

Soit x_{ij} la valeur prise par la variable x sur l'unité (ij) .

Supposons qu'on ait : $1 \leq i \leq \mu$; $1 \leq j \leq \bar{\nu}$ et qu'on tire au sort m unités primaires (sur μ), puis $\bar{n}$ unités du second degré dans chaque unité primaire.

On sait alors que la moyenne générale des unités du 2ème degré

$$\bar{\bar{x}} = \sum_i \sum_j x_{ij} / \mu \bar{\nu}$$

admet comme estimation sans biais

$$\bar{X} = S_i S_j x_{ij} / m \bar{n}$$

dont la variance est :

$$\forall \bar{X} = \sum_i \frac{1}{m\mu} \frac{\bar{\nu} \sigma_i^2}{\bar{\nu} - 1} \left(\frac{1}{\bar{n}} - \frac{1}{\bar{\nu}} \right) + \frac{\mu \sigma_0^2}{\mu - 1} \left(\frac{1}{m} - \frac{1}{\mu} \right)$$

avec

$$\sigma_i^2 = \mathcal{V}(x_{ij}/i); \quad \bar{x}_i = \sum_j x_{ij}/\bar{v}$$

$$\sigma_o^2 = \mathcal{V}(\bar{x}_i).$$

Supposons que cette variance représente une perte d'information. Considérons la population formée par l'échantillon de taille $(m, \bar{n})$. Tirons-en un échantillon de taille $(m', \bar{n}')$.

$$m' < m, \quad \bar{n}' < \bar{n}$$

La perte d'information supplémentaire serait :

$$\frac{1}{m'm} \sum_i \frac{\bar{n} s_i^2}{\bar{n} - 1} \left(\frac{1}{\bar{n}'} - \frac{1}{\bar{n}} \right) + \frac{m s_o^2}{m - 1} \left(\frac{1}{m'} - \frac{1}{m} \right)$$

Prenons l'espérance mathématique de cette expression pour tous les échantillons de taille $(m, \bar{n})$. On a :

$$E_2(\bar{n} s_i^2 / \bar{n} - 1) = \bar{v} \sigma_i^2 / \bar{v} - 1; \quad \mathcal{E}(m s_o^2 / m - 1) = \mu \sigma_o^2 / \mu - 1$$

$$\mathcal{E} \left[\frac{1}{m'm} \sum_i \frac{\bar{v} \sigma_i^2}{\bar{v} - 1} \right] = \frac{1}{m'\mu} \sum_i \frac{\bar{v} \sigma_i^2}{\bar{v} - 1}$$

Ainsi la nouvelle perte (totale) d'information serait :

$$\frac{1}{\mu} \sum_i \frac{\bar{v} \sigma_i^2}{\bar{v} - 1} \left[\frac{1}{m'} \left(\frac{1}{\bar{n}'} - \frac{1}{\bar{n}} \right) + \frac{1}{m} \left(\frac{1}{\bar{n}} - \frac{1}{\bar{v}} \right) \right] + \frac{\mu \sigma_o^2}{\mu - 1} \left(\frac{1}{m'} - \frac{1}{m} \right)$$

au lieu de :

$$\frac{1}{\mu} \sum_i \frac{\bar{v} \sigma_i^2}{\bar{v} - 1} \frac{1}{m'} \left(\frac{1}{\bar{n}'} - \frac{1}{\bar{v}} \right) + \frac{\mu \sigma_o^2}{\mu - 1} \left(\frac{1}{m'} - \frac{1}{m} \right)$$

Conclusion.

Quand on change simultanément le nombre des unités primaires échantillon et celui des unités du 2ème degré, la variance $\bar{X}$ ne serait plus une perte d'information.

Le vice de raisonnement est dans l'expression supposée de la perte d'information supplémentaire, qui découle d'une interprétation trop simpliste de la transformation

$$(\zeta Z) \longrightarrow (Z_1 Z_2)$$

L'interprétation correcte de (ζZ) deviendra très naturelle après les considérations géométriques qu'on trouvera au § II.

C - REMARQUE - LA THEORIE `DE BASU.

Basu (Ba 1, 1952; Ba 2, 1954) nous a précédé, et notre indicateur (ζ Z) est proche parent de sa fonction de risque $\mathcal{E}W$. Nous divergerons vite, car :

a) nous avons cherché quelles fonctions de risque étaient solution de la relation de Chasles, devenue équation fonctionnelle (de Chasles);

b) nous avons recours à des considérations géométriques;

c) finalement nous sommes obligé d'élargir la définition de Basu pour définir la perte d'information dans des cas pourtant fort courants.

Basu a tels soucis de rigueur et de généralité que nous citerons in extenso le début de son premier article.

"Soit $X_1 X_2 \dots X_n$ un système de variables aléatoires dont on sait que la fonction de distribution $F(x_1 x_2 \dots x_n)$ appartient à un certain sous-espace Ω de l'espace de toutes les fonctions de distribution possibles F . Par exemple, il arrivera qu'on sache que les X sont des distributions identiques mais indépendantes, de sorte que Ω est le sous-espace de toutes les fonctions de distribution de la forme :

$$F = G(x_1) \cdot G(x_2) \dots G(x_n) \quad (1.1)$$

"G étant une quelconque fonction de distribution d'une seule variable.

"Le problème de l'estimation ponctuelle est celui de l'estimation d'une certaine caractéristique $\theta = \mu(F)$ de la population, [$\mu(F)$ étant une fonctionnelle à valeurs réelles définie pour tout $F \in \Omega$] à l'aide d'un estimateur $t = t(x_1 x_2 \dots x_n)$ où x_i est une observation aléatoire de la variable X_i .

"Désignons par $W(t, F)$ pour tout $R \in \Omega$ fixé, les divers poids que le statisticien attache aux différentes valeurs de t en tant qu'estimations de $\mu(F)$; et soit

$$r(F/t) = \int_R W(t, F) dF \quad (1.2)$$

"(où R est l'espace échantillon à n dimensions) la fonction de risque associée à l'estimateur t .

"Nous supposons qu'il existe des estimateurs t pour lesquels l'intégrale (1.2) est convergent pour tout $F \in \Omega$.

"Si $r(F/t_1) \leq r(F/t_2)$ pour tout $F \in \Omega$, le signe $=$ étant valable pour au moins un F , alors on dit que t_1 est uniformément plus puissant que t_2 .

"L'estimateur t_0 peut être dit admissible s'il n'existe aucun estimateur t uniformément plus puissant que t_0 .

"Dans le présent article nous nous limiterons aux seules fonctions poids qui sont des fonctions convexes (en décroissant) de t pour tout $F \in \Omega$ c'est-à-dire $2W(t_1 + t_2/2, F) \leq W(t_1, F) + W(t_2, F)$ pour tous les t_1 et t_2 . Si le signe $=$ ne vaut que pour $t_1 = t_2$, on dira que la fonction est strictement convexe".

Tout ceci reste valable ici, mais pour un autre usage. Basu se donne a priori une forme d'estimateur (par exemple linéaire pour estimer la moyenne, quadratique pour estimer la variance); et il se donne aussi l'expression du "risque" élémentaire W . Considérant alors le "risque" $r = \mathcal{E}W$, il détermine les paramètres des estimateurs de telle sorte que le "risque" soit minimum. Comme nous l'avons dit ailleurs (Réf. T. 11, 1956) lorsqu'on tire des boules d'une urne, avec

$$W = (Z - \zeta)^2,$$

le risque étant $r = \mathcal{E}(Z - \zeta)^2 = \mathcal{V}Z + b^2$,

(où b désigne le biais), la politique du moindre risque ne conduit pas du tout à prendre $b = 0$ puis $\mathcal{V}Z$ minimum; elle conduit à prendre des estimateurs biaisés.

Par exemple, avec :

$$\zeta = \bar{x}, \quad Z = c\bar{X}, \quad \mathcal{V}Z = c^2\mathcal{V}\bar{X}, \quad \gamma^2 = \mathcal{V}\bar{X}/\bar{x}^2,$$

le moindre risque ne correspond pas à $c = 1$, mais à

$$c = 1/(1 + \gamma^2)$$

Lorsque γ^2 est inconnu, Basu (suivant les méthodes de Wald) définit tout l'intervalle $0 < c < 1$ comme fournissant les estimateurs uniformément les plus puissants.

L'intérêt du cas banal $c = 1$ n'apparaît que si l'on se souvient qu'en pratique γ^2 est très petit, donc c très peu inférieur à 1.

Ce résultat troublant disparaît si, au lieu de définir ainsi le "risque" trop arbitrairement, on lui impose d'être une perte d'information - au sens où Schutzenberger entend information.

Alors $\mathcal{E}(Z - \zeta)^2$ par exemple n'est une perte d'information qu'avec un estimateur sans biais (uniformément); $\mathcal{E}Z^2 - \zeta^2$ serait une perte d'information (sous des conditions plus larges) mais ce n'est plus un "risque" au sens de Wald, c'est-à-dire qu'on n'a pas en général

$$\mathcal{E}Z^2 - \zeta^2 = \mathcal{E}W \geq 0$$

(avec $W \geq 0$). Les estimations biaisées amènent donc une rupture entre le point de vue de Basu et le nôtre (voir Chapitres IV et VII).

II - TOPOLOGIE DES SONDAGES -

A - L'ESPACE TOPOLOGIQUE.

L'intérêt (voire la nécessité) d'introduire des considérations de topologie dans les sondages n'apparaît que quand on dépasse le stade de l'urne unique, un peu avec le sondage stratifié, beaucoup avec le sondage à plusieurs degrés. On évoquera ici quelques treillis sans recourir à aucun théorème de topologie. On a seulement voulu préciser où finissait le besoin d'un espace topologique, où commençait celui d'une métrique.

1/ - Figuration des échantillons au moyen d'un arbre⁽¹⁾.

Ce n'est pas que le procédé classique (en théorie des Communications) de figuration par un arbre, d'un phénomène aléatoire ne soit applicable aux sondages, - mais il devient vite assez lourd. Décrivons-le cependant dans le cas le plus simple :

Exemple : Cas des tirages sans remise dans une urne d'effectif v .

Un point figure la connaissance complète de l'urne (urne vidée). De ce point partent v branches aboutissant à v points dont chacun représente un échantillon d'effectif $v - 1$.

- chaque branche se ramifie en $v - 1$ sous-branches, qui se rejoignent deux par deux, en $v(v - 1)/2$ points représentant les échantillons d'effectif $v - 2$;

- chaque sous-branche se ramifie en $v - 2$ rameaux convergeant trois par trois, etc.

On arrive ainsi à C_v^{v-2} point (sondages d'effectif $n = 2$), d'où partent deux rameaux; ceux-ci se rejoignent par $(v - 1)$ à la fois en v points figurant les échantillons d'effectif $n = 1$.

(1) En topologie, "arbre" a un sens moins large qu'ici; "graphe" serait trop général.

On observera que les échantillons d'effectif 0 ne figurent pas sur l'arbre; ils n'apportent aucune "information"; mais il arrivera que les échantillons d'effectif 1, 2, ... ne renseignent pas non plus sur tel aspect de la population sondée qui justement nous intéresse; par exemple un échantillon d'effectif $n = 1$, ne permet pas d'estimer la variance.

Remarque.

On peut également définir "l'arbre" dans l'autre sens, à partir de v échantillons d'effectif 1, - donnant $v(v-1)/2$ échantillons d'effectif 2, etc. si les tirages sont exhaustifs, - ou v^2 échantillons d'effectif 2, etc. si les tirages sont bernoulliens (auquel cas l'arbre ne se referme jamais).

2/ - Figuration (topologique) des tirages dans une urne.

On va symboliser l'arbre du n°1 ci-dessus sur un axe d'information.

Soit v l'effectif de l'urne, n l'effectif de l'échantillon; on peut représenter par une suite de points consécutifs (n) sur un demi-axe, les sondages d'effectif 1, 2, ... n ...

Le point (v) représentant la connaissance parfaite de l'urne sera pris comme origine du demi-axe.

Il peut être commode de distinguer un point (O) représentant l'absence d'échantillon. On peut le supposer rejeté à l'infini.

En faisant tendre v vers l'infini, on figure le cas d'une population de base infinie. Les tirages avec remise seront figurés de même. Enfin il n'a pas été spécifié que les tirages étaient équiprobables ou non, et les tirages (bernoulliens) à probabilités inégales sont figurés par le même axe.

3/ - Représentation topologique d'un plan de sondage.

Etant donné la population découpée en strates, unités de sondages, etc., autrement dit une structure de sondage, un plan de sondage appartenant à cette famille de sondages se définit en fixant un certain nombre de paramètres : soit effectifs de l'échantillon, soit fractions de sondage. Chaque plan de sondage de la famille étant représenté par un point d'un espace topologique, ce point occupe les divers sommets d'un treillis lorsqu'on donne aux paramètres toutes les valeurs admissibles.

Quant aux échantillons eux-mêmes, ils seraient représentés par l'arbre; mais celui-ci sera refermé sur son axe, chaque point ou som-

met du treillis figurant tous les échantillons qu'un même plan de sondage peut procurer.

Une fois définis les points de cet espace, les autres éléments essentiels en sont les droites ou "géodésiques". Mais par deux points du treillis ne passe pas nécessairement une droite (et une seule); et nous préférons axes ou génératrices à géodésiques.

L'existence d'un "côté" ou segment entre deux points du treillis est un fait concret; elle signifie qu'on passe du plan de sondage π au plan de sondage π' par une opération matérielle de tirage au sort (per-tes d'une ou plusieurs unités de sondage). En revanche le problème est de savoir si 3 points (et davantage) du treillis sont ou non alignés sur une même génératrice.

Il est naturel de le résoudre en introduisant sur le treillis une métrique, qui est l'"information"⁽¹⁾

La distance entre deux points étant ainsi définie, l'alignement des 3 points A B C s'obtient en écrivant

$$AB + BC = AC$$

Mais on peut objecter que les axes du treillis dépendent alors de la métrique et peuvent changer quand on change de métrique.

Une autre difficulté est l'existence de certain triangles ABC sur le treillis, la distance AC étant alors égale à $AB + BC$ sans qu'il y ait alignement.

Exemple de triangle sur le treillis.

Considérons un sondage stratifié avec deux strates d'effectifs (v_1, v_2). Par exemple $v_1 = 20, v_2 = 100$. Au grillage relatif au sondage (n_1, n_2) se superpose un demi-axe relatif aux 20 sondages à fractions sondées égales f , soit :

$$f = 1/20, 1/10, 3/20, 1/5, 1/4, \text{ etc.}$$

$$n = n_1 + n_2 = 6, 12, 18, 24, 30, 36, \text{ etc.}$$

concernant les couples (n_1, n_2) suivants :

(1) L'information est une "valuation de treillis", Schützenberger, Thèse, p. 40.

(1, 5) (2, 10) (3, 15) (4, 20) (5, 25) (6, 30) etc.

On a donc par exemple le "triangle" suivant :

$$(3, 15) = (\pi); (3, 10) = (\pi'); (2, 10) = (\pi'')$$

C'est pourquoi on prendra les précautions suivantes :

a) Définition des arêtes telle qu'il n'existe plus de triangles.

On figurera par une arête du treillis tout tirage dans une seule urne, opération faisant passer d'un sommet à un autre, d'un plan de sondage à un autre de taille inférieure. Il est clair qu'on peut toujours (et au moins d'une façon) cheminer sur les arêtes, de l'origine (point figurant la population prise en totalité) à tout sommet du treillis.

b) Définition d'axes d'information invariants si la métrique change.

Par définition sont alignés les points représentant les plans de sondage ($v, v-1, v-2, v-3 \dots$) tels qu'on expulse d'une urne donnée (0, 1, 2, 3, ...) unités de sondage, toutes choses égales d'ailleurs.

Ceci revient à dire que, quelle que soit la perte d'information, les plans de sondage, qui sont des tirages dans une urne unique, sont figurés par un axe.

Les plans de sondage (si compliqués soient-ils) étant des combinaisons de tirages dans une urne, on possède donc une définition "opérationnelle" de certains axes du treillis; l'existence d'autres axes reste possible.

4/ - Exemples de treillis de sondage.

a) Cas des sondages stratifiés (avec 2 strates seulement et 2 paramètres).

Le grillage (n_1, n_2) représente (on l'a vu) un sondage à 2 strates (fractions sondées distinctes); le point $U(v_1, v_2)$ représente l'état de connaissance parfaite des urnes. Les points représentant les sondages à fractions sondées égales f auraient en outre été alignés sur un demi-axe d'origine U ; mais il s'agit en fait d'un deuxième treillis distinct du premier, correspondant à un estimateur de x dépendant d'un seul paramètre f .

En convenant que les arêtes n'existent qu'avec les pertes d'une boule à la fois, nous perdons le deuxième treillis et les triangles qu'il dessinait avec le grillage.

La convention faite sur les axes nous montre que tous les "barreaux" du grillage sont axes d'information.

b) Cas des sondages à 2 degrés et 2 paramètres.

Soit un univers où les μ unités du 1er degré comprennent des nombres égaux d'unités du 2ème degré : l'échantillon est défini par le nombre m d'unités du 1er degré tirées et le nombre $\bar{n}$ d'unités du 2ème degré extraites de chacune des précédentes; on le représentera par les sommets du grillage ($m, \bar{n}$).

Le point $(\mu, \bar{v})$ correspond à une connaissance parfaite; l'axe $(m, \bar{v})$ représente un sondage en grappes et l'axe $(\mu, \bar{n})$ un sondage stratifié (à fractions sondées égales), ce sont deux axes d'information.

La définition opérationnelle des axes nous donne les axes (m constant, $\bar{n}$ variable) (1er système).

Il existe un second procédé de sondage à 2 degrés qui consiste à tirer d'abord $\bar{n}$ unités secondaires dans chaque unité primaire, puis à désigner au sort m sur μ unités primaires (2^e système d'axes).

Les deux procédés sont équivalents parce que ($\bar{n}$ étant constant) les tirages du 1er et du 2ème degrés sont indépendants. On dispose donc ici encore de deux systèmes d'axes d'information croisés.

c) Cas d'un sondage à 2 phases (et 2 paramètres).

On tire d'abord m unités (avec ou sans remise) d'une urne d'effectif μ . On les trie en trois strates d'effectifs $m/3$ et on tire $\bar{n}$ unités de chaque strate. Le treillis formé des points $(m, \bar{n})$ admet, d'une part, l'axe $(m, m/3)$, d'autre part, les axes (m constant, $\bar{n}$ variable). Il n'existe pas de second système d'axes croisé avec le premier.

d) Cas d'un sondage à 2 degrés avec dépendance du 2ème degré.

Si l'on sonde un univers où les unités primaires sont de tailles très inégales, on a parfois intérêt à se donner l'effectif total n au 2ème degré, à répartir entre les m unités du 1er degré échantillon, au prorata de leurs tailles v_i . (C'est un cas d'estimation biaisée).

On a ainsi $n_i/n = v_i/S v_i$ et le 2ème degré de sondage dépend stochastiquement du 1er.

On ne peut tirer d'abord les unités du 2ème degré puis celles du 1er degré. Il n'existe qu'un système d'axes d'information, s'appuyant sur un axe unique transversal.

5/ - Opérations sur les treillis de sondage.

En passant en revue les problèmes courants de sondage, on constate qu'il est souvent maladroit de représenter directement les plans de sondage de paramètres $f_1 f_2 f_3 \dots$ par le treillis $(f_1, f_2, f_3 \dots)$. Il peut être plus commode de considérer un treillis plus vaste : soit qu'on en découpe une partie seule utile, soit qu'on y figure les échantillons possibles avec leurs probabilités respectives, le treillis jouant alors un rôle analogue à celui de l'arbre défini au début.

Plus généralement, le treillis primitif sera découpé en un sous-ensemble des points exclus et divers autres sous-ensembles dont chacun représente un nouveau plan de sondage.

Donnons des exemples.

a) Sondage stratifié avec fractions de sondage égales f.

Du treillis $(n_1, n_2, \dots, n_h)$ on conserve les points $(fv_1, fv_2, \dots, fv_h)$, sous-ensemble décrit quand f (ou $n = fv$) varie.

b) Sondage stratifié "optimum".

Du même treillis on conserve les points $\left(\frac{\lambda v_1 \sigma_1}{\sqrt{c_1}}, \frac{\lambda v_2 \sigma_2}{\sqrt{c_2}}, \dots, \frac{\lambda v_h \sigma_h}{\sqrt{c_h}} \right)$ sous-ensemble décrit quand λ (ou n) varie (avec $n = \lambda \sum v_h \sigma_h / \sqrt{c_h}$).

c) Stratification a posteriori.

Le même treillis est découpé en sous-ensembles tels que $\sum n_h = n$.

Chaque point d'un sous-ensemble donné est affecté d'une probabilité hypergéométrique (tirage exhaustif) ou multinomiale (tirage bernoullien).

Toutefois, on aura pris soin d'éviter qu'aucun des points (n_h) ne corresponde à une perte d'information infinie introduite artificiellement par un estimateur inadéquat. L'apparition de certains n_h nuls est un évènement de probabilité non nulle; et avant tirage de l'échantillon il faut convenir de l'estimation à employer si cela se produit.

B - PASSAGE DE L'ESPACE TOPOLOGIQUE A L'ESPACE METRIQUE.

1/ - Perte d'information.

On attache à présent à chaque sommet du treillis une expression p - la perte d'information -

- nulle au sommet U représentant la connaissance parfaite de l'univers sondé;
- infinie aux points représentant un état d'absence complète de connaissance;
- satisfaisant à la relation de récurrence, à savoir :

$$p(\pi') = p(\pi) + \mathcal{E} P(\pi' / \pi)$$

- en désignant par π et π' 2 plans de sondages représentés par 2 points (ou sommets) du treillis non pas quelconques mais reliés par une arête (ou côté);

- et par $P(\pi' / \pi)$ une expression qui se réduirait à $p(\pi')$ si l'échantillon du plan de sondage (π) recouvrait toute la population; autrement dit :

$$P(\pi' / U) \equiv p(\pi')$$

En employant une majuscule P, nous entendons spécifier qu'il s'agit d'une grandeur aléatoire.

Quand on écrit $\mathcal{E} P(\pi' / \pi)$, on désigne l'espérance mathématique de $P(\pi' / \pi)$ étendue à tous les échantillons du plan de sondage (π).

Partant du point U, et cheminant exclusivement le long des côtés du treillis, on obtient finalement l'information perdue par chaque plan de sondage de la famille; lorsque (π) peut être atteint par plusieurs itinéraires, la fonction de treillis $p(\pi)$ a (par hypothèse) une valeur indépendante de l'itinéraire (à condition qu'il s'agisse bien du même plan de sondage π).

2/ - Métrie.

Simultanément on introduit une métrique dans cet espace, en définissant la distance entre deux points situés sur une même géodésique comme égale à la variation de la perte d'information.

$$(\text{distance } \pi \pi') = p(\pi') - p(\pi) = \mathcal{E} P(\pi' / \pi)$$

3/ - Conséquence.

Relation entre les informations perdues par 3 plans π π' π'' de la famille de plans de sondage considérée :

Désignant par (π) (π') (π'') 3 plans de sondage de la famille considérée, représentés par trois sommets du treillis reliés deux à deux par des côtés du treillis,

$$\text{on a} \quad p(\pi') = p(\pi) + \mathcal{E}P(\pi'/\pi)$$

$$p(\pi'') = p(\pi') + \mathcal{E}P(\pi''/\pi')$$

$$\text{donc} \quad p(\pi'') = p(\pi) + \mathcal{E} \left[P(\pi''/\pi') + P(\pi'/\pi) \right]$$

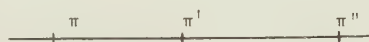
Mais si π'' et π sont également sur un côté, on a :

$$p(\pi') - p(\pi) = \mathcal{E}P(\pi'/\pi)$$

D'où :

$$\mathcal{E}P(\pi''/\pi) = \mathcal{E}P(\pi''/\pi') + \mathcal{E}P(\pi'/\pi)$$

Cette condition signifie (en l'absence de triangles sur le treillis) que $\pi \pi' \pi''$ sont alignés sur un axe d'information portant à la fois le côté $\pi \pi'$, le côté $\pi' \pi''$ et le côté $\pi \pi''$.



On sait que cette relation doit subsister quel que soit P .

4/ - Relation entre $p(\pi)$ et $p(\pi'')$.

π et π'' étant quelconques, on aura :

$$p(\pi'') = p(\pi) + \mathcal{E}P(\pi'_1/\pi) + \mathcal{E}_1 P(\pi'_2/\pi'_1) + \dots + \mathcal{E}_k P(\pi''/\pi'_k)$$

à condition qu'on puisse s'arranger pour que $\pi \pi'_1 \pi'_2 \dots \pi'_k \pi''$ forme une ligne brisée composée de portions d'axe d'information.

Application.

De la famille F des plans (π) représentée par le treillis, extrayons une suite de plans de sondage 1, 2, 3, ... n, telle que "la taille de l'échantillon décroisse constamment" et que $Z_i Z_{i+1}$ (consécutifs) soient toujours sur un même axe d'information. On a par conséquent :

$$(\zeta Z_1) + \mathcal{E}_1(Z_1 Z_2) + \mathcal{E}_2 \mathcal{E}_1(Z_2 Z_3) + \dots \mathcal{E}_1 \dots \mathcal{E}_{n-1}(Z_{n-1} Z_n) = (\zeta Z_n)$$

En posant :

$$\mathcal{E}_1 \dots \mathcal{E}_i(Z_i Z_{i+1}) = \overline{Z_i Z_{i+1}}$$

la formule s'écrit :

$$\overline{\zeta Z_1} + \overline{Z_1 Z_2} + \overline{Z_2 Z_3} + \dots + \overline{Z_{n-1} Z_n} = \overline{\zeta Z_n}$$

C'est la formule de Chasles (généralisée).

Conventions.

a) Jusqu'ici d'après nos conventions, les termes $\overline{\zeta Z_1}$, $\overline{Z_1 Z_2}$, etc. sont positifs.

Pour pouvoir parler de relation de Chasles, il est nécessaire de définir aussi les termes négatifs. On va donc convenir que :

$$\text{si } \overline{Z_i Z_j} \text{ est négatif, il est égal à } - \overline{Z_j Z_i}$$

où $\overline{Z_j Z_i}$ (positif) désigne $\mathcal{E}_1 \dots \mathcal{E}_j (Z_j Z_i)$.

b) Convention de langage.

On appellera $\overline{Z_j Z_i}$ (positif) perte d'information quand on passe de (j) à (i) et gain d'information quand on passe de (i) à (j). En particulier (ζZ_1) est le gain d'information quand on passe du plan d'échantillonnage (1) à une connaissance complète de la population.

Conséquences.

a) La "formule de Chasles" s'étend formellement à tous les plans de sondages de F, sans qu'il soit encore nécessaire de supposer : les tailles décroissantes, les plans de sondage consécutifs placés sur une même géodésique ; mais les symboles $\overline{Z_i Z_1}$ perdent leur signification primitive.

b) Le long de tout circuit fermé, s'il en existe sur les axes du treillis de F, on obtiendra :

$$\overline{Z_1 Z_2} + \overline{Z_2 Z_3} + \dots + \overline{Z_k Z_1} = 0$$

Conclusion.

La perte d'information apparaît comme la somme $\sum X_n$ de composantes dont chacune est la "longueur" d'une portion d'axe du treillis.

En étudiant des plans de sondage concrets, on découvrira les conséquences de l'existence de circuits sur le treillis (sondage à plusieurs degrés) et on s'apercevra de l'existence de composantes sous-tractives (sondage à plusieurs phases).

Mais, au préalable, on va chercher (au Ch. IV) à résoudre l'équation fonctionnelle de Chasles, c'est-à-dire à connaître la forme analytique des P.

Remarque. Les changements de découpage et l'information.

Il est en pratique très fréquent que les ζ à estimer soient indépendants du découpage : c'est, par exemple, le cas si on veut estimer la superficie en blé, en France.

La perte d'information est définie à l'aide d'un treillis, qui traduit une certaine structure de sondage. Le treillis se conserve si l'on garde la structure tout en modifiant le découpage du territoire en unités, en strates, ... ; les valeurs de la perte d'information et la représentation métrique du sondage se modifient; ceci traduit les changements internes de l'estimateur (apparemment immuable), pour l'estimation d'un paramètre qui, lui, est invariant dans les transformations du découpage.

Supposons à présent qu'au lieu de la superficie en blé on veuille estimer la variance ξ entre unités primaires de sondage; ξ n'est pas invariante quand on modifie le découpage du territoire en unités de sondage.

On estime simultanément: la superficie en blé ζ (à l'aide de Z), et la perte d'information $\mathcal{V}Z$ (à l'aide d'un certain W), en supposant Z sans biais. L'expression de $\mathcal{V}Z$ renferme ξ . On s'intéresse parfois à la perte d'information relative à W (c'est-à-dire une perte d'ordre 2). Lorsque le découpage est modifié, les variations de cette dernière perte ne signifient plus rien: car $\mathcal{V}Z$ et W variant simultanément, leur "distance" se modifie sous une double cause. Z estime sans biais ζ malgré le découpage changé, alors qu'on ne sait plus trop quel $\mathcal{V}Z$ est estimé par W .

Nous pensons donc que la théorie de la perte d'information doit être faite avec un ou plusieurs découpages bien déterminés, qu'elle ne supporte pas les changements de découpage.

Application aux sondages à plusieurs degrés.

Empruntons à une étude encore inédite de M. Fonsagrive un cas intéressant de sondages à plusieurs degrés. On suppose que (comme c'est souvent le cas en pays sous-développé) le découpage des unités du 2ème degré et la sous-stratification n'ont lieu que dans les unités échantillon du 1er degré, en fonction des singularités que présentent ces dernières et en vue de réduire le plus possible la variance (à l'intérieur de l'échantillon du 1er degré); ainsi la vraie structure de sondage reste-t-elle en grande partie indéterminée.

Sans doute est-il encore permis de parler de l'ensemble des découpages possibles, et de concevoir un estimateur sans biais, sa va-

riance et l'estimateur sans biais de cette variance pour chacun de ces découpages; mais l'expérience pratique se limite à un seul découpage, en partie embryonnaire, et un seul échantillon; et la variance ainsi estimée manque de signification précise vis-à-vis de toute la partie non spécifiée du découpage.

Autres applications.

Avec les procédés courants de stratification a posteriori ou de sondage à deux phases, on rencontre la même difficulté : le découpage en strates n'est arrêté qu'après tirage de l'échantillon, ou du grand échantillon. Si ledit tirage avait donné un autre résultat, on aurait modifié les strates et, par conséquent, la variance d'échantillonnage. L'estimation de cette variance, sur l'échantillon effectivement tiré, a donc une signification limitée.

Pour qu'il en soit autrement, il faudrait (pour éviter toute difficulté) se donner à l'avance des règles prévoyant (dans tous les cas possibles) le découpage en strates à effectuer et l'estimateur à employer.

Nous nous en tiendrons ici à des plans de sondage déterminés par un nombre fini de paramètres, c'est-à-dire à des cadres rigides de sondage.

CHAPITRE IV

LA FORME ANALYTIQUE DES PERTES D'INFORMATION

On va rechercher quelles formes les expressions $p(\pi)$ et $P(\pi'/\pi)$ sont susceptibles de prendre lorsqu'on ne sait absolument rien de la forme de la distribution sondée, hypothèse très réaliste en théorie des sondages.

I - PERTES AU SENS DU CHAPITRE I -

1/ - Introduction.

Au sens du Chapitre I, la perte d'information concerne plus spécialement l'estimation d'une certaine caractéristique ζ par un estimateur Z , supposé sans biais (appréciable).

On va supposer la perte d'information de la forme $\mathcal{E} W(Z)$ c'est-à-dire fonction de risque (au sens de Wald); cherchons quelles conditions doit satisfaire W pour que la fonction de risque $\mathcal{E} W(Z)$ soit perte d'information.

Par hypothèse on a $W(\zeta) = 0$, $W(Z) > 0$ si $Z \neq \zeta$.

On supposera d'abord W fonction de Z ; on sera amené ensuite à considérer W comme fonctionnelle de Z . Cette fonction de Z sera supposée convexe; mais nous tenons à conserver le cas très important où $W = |Z - \zeta|$; il ne faut donc pas supposer W "strictement convexe".

Le problème est à présent de trouver W tel que $\mathcal{E} W$ satisfasse à l'équation de Chasles.

Il existe des solutions évidentes de cette équation fonctionnelle qui peut s'écrire :

$$\text{avec } \begin{cases} \mathcal{E} W(Z_1, \zeta) = \sum_h p_1^h W(z_1^h, \zeta) \\ \mathcal{E} W(Z_2, \zeta) = \sum_k p_2^k W(z_2^k, \zeta) \end{cases}$$

$$\sum_k p_2^k W(z_2^k, \zeta) = \sum_h p_1^h W(z_1^h, \zeta) + \sum_h p_1^h \sum_k p_2^k W(z_2^k, z_1^h)$$

2/ - Une solution particulière symétrique de l'équation de Chasles.

Ne retenons que les cas où il est possible d'intervertir z_1^h et z_2^k dans les formules.

En effet il serait gênant qu'on ait :

$$\sum_h \sum_k p_1^h p_2^k W(z_1^h, z_2^k) = \overline{Z_1 Z_2}$$

alors que $\sum_k \sum_h p_2^k p_1^h W(z_2^k, z_1^h)$ aurait une autre signification. Ainsi le cas particulier où $W(z, \zeta)$ est symétrique semble-t-il présenter le plus d'intérêt.

Les exemples de fonction W donnés par Basu sont justement des fonctions symétriques en ζ et Z .

On a alors par définition :

$$\overline{Z_1 Z_2} = \begin{cases} \mathcal{E}W(Z_2, Z_1) & \text{si Taille (1) > Taille (2)} \\ \mathcal{E}W(Z_1, Z_2) & \text{si Taille (2) > Taille (1)} \end{cases}$$

(le cas où les tailles ne sont pas directement comparables est exclu).

a) L'équation de Chasles

$$\mathcal{E}W(Z_2, \zeta) = \mathcal{E}W(Z_1, \zeta) + \mathcal{E}W(Z_2, Z_1)$$

admet comme solution particulière :

$$W(Z, \zeta) = (Z - \zeta)^2$$

à condition que :

$$\mathcal{E}[(\zeta Z_1)(Z_1 Z_2)] = 0 ;$$

: dans l'espace (A) où on représente l'aléatoire Z_1 par un point, les vecteurs

$$\overrightarrow{\zeta Z_1} \text{ et } \overrightarrow{Z_1 E Z_2}$$

doivent donc être orthogonaux; à moins qu'on n'ait (estimateur uniformément sans biais)

$$E Z_2 = Z_1 \quad (\text{donc } \mathcal{E} Z_1 = \zeta)$$

b) On étend le résultat à :

$$W(Z, \zeta) = [g(z) - g(\zeta)]^2$$

où $g(z)$ est une fonction quelconque; toujours sous réserve qu'on ait :

$$\mathcal{E}\{[g(Z_1) - g(\zeta)] [g(Z_2) - g(Z_1)]\} = 0$$

3/ - Une solution non symétrique de l'équation de Chasles.

L'absence de biais de l'estimateur Z correspond à la condition

$$2 \mathcal{E}(\zeta - Z) = 0$$

c'est-à-dire $\frac{\partial}{\partial \zeta} \mathcal{E}(\zeta - Z)^2 = 0$, c'est-à-dire $\frac{\partial}{\partial \zeta} \mathcal{E}W = 0$ avec $W = (\zeta - Z)^2$.

On peut écrire aussi

$$\mathcal{E}W = \mathcal{E}Z^2 - 2\zeta \mathcal{E}Z + \zeta^2 = \underline{\mathcal{E}Z^2 - \zeta^2}$$

Sous cette dernière forme, il est clair que la relation de récurrence sera vérifiée.

De même l'absence de biais peut s'écrire encore :

$$\mathcal{E}[\varphi(\zeta)(\zeta - Z)] = 0$$

Posons $W(Z, \zeta) = \int \varphi(\zeta)(\zeta - Z)d\zeta$ avec

$$\int \dot{\varphi}(\zeta)d\zeta = \varphi_1(\zeta); \quad \int \zeta \dot{\varphi}(\zeta)d\zeta = \varphi_2(\zeta)$$

$$W = \varphi_2(\zeta) - \varphi_2(Z) - Z [\varphi_1(\zeta) - \varphi_1(Z)]$$

$$\mathcal{E}W = \mathcal{E}[Z\varphi_1(Z) - \varphi_2(Z)] - [\zeta\varphi_1(\zeta) - \varphi_2(\zeta)]$$

Là encore il apparaît que la relation de récurrence sera vérifiée.

1ere application : $\varphi(\zeta) = 2\zeta$ avec $\zeta > 0$,

$$\text{d'où} \quad W = (\zeta - Z)^2 \left(\zeta - \frac{\zeta - Z}{3} \right) = \frac{1}{3} (2\zeta^3 - 3\zeta^2 Z + Z^3)$$

$$\mathcal{E}W = \frac{1}{3} (2\zeta^3 - 3\zeta^2 \mathcal{E}Z + \mathcal{E}Z^3)$$

$$\mathcal{E}W \equiv \frac{1}{3} (\mathcal{E}Z^3 - \zeta^3)$$

2ème application. $\varphi(\zeta) = \zeta^2$

$$\text{d'où} \quad W = (\zeta - Z)^2 \zeta^2 - \frac{1}{2} (\zeta - Z)^4 - \frac{2Z}{3} (\zeta - Z)^3$$

$$= \frac{1}{4} \zeta^4 - \frac{2}{3} \zeta^3 Z + \frac{1}{6} Z^4$$

$$\mathcal{E}W \equiv \frac{1}{6} (\mathcal{E}Z^4 - \zeta^4)$$

Ces exemples mettent sur la voie de résultats plus généraux.

4/ - THEOREME I -

$\mathcal{E}W$ est nécessairement de la forme $\mathcal{E}\phi(Z) - \phi(\zeta)$, (où ϕ est une fonction), que l'estimateur Z soit sans biais ou qu'il soit biaisé, - du moment de W est fonction de ζ .

En effet partons de la relation de récurrence des pertes d'information :

$$(1) \quad \mathcal{E}W(Z_2, \zeta) - \mathcal{E}W(Z_1, \zeta) = \mathcal{E}EW(Z_2, Z_1).$$

On voit que le second membre ne dépend pas de ζ ; il en résulte que $\mathcal{E}W(Z_1, \zeta) = f_1(\zeta)$ et $\mathcal{E}W(Z_2, \zeta) = f_2(\zeta)$ sont nécessairement de la forme $k_1 - \phi(\zeta)$ et $k_2 - \phi(\zeta)$, $\phi(\zeta)$ s'éliminant par différence.

Si l'on substitue Z_1 à ζ dans $W(Z_2, \zeta)$, il suit de là que

$$\mathcal{E}W(Z_2, \zeta) = k_2 - \phi(\zeta)$$

$$\text{entraîne} \quad EW(Z_2, Z_1) = K_{21} - \phi(Z_1)$$

$$\text{tandis que} \quad \mathcal{E}W(Z_1, \zeta) = \mathcal{E}W(Z_2, \zeta) - \mathcal{E}EW(Z_2, Z_1)$$

$$\text{implique} \quad \mathcal{E}W(Z_1, \zeta) = k_2 - \phi(\zeta) - [\mathcal{E}K_{21} - \mathcal{E}\phi(Z_1)]$$

$$\text{ou} \quad \mathcal{E}W(Z_1, \zeta) = [k_2 - \mathcal{E}K_{21}] + \mathcal{E}\phi(Z_1) - \phi(\zeta)$$

où le crochet comprend deux termes dépendant du sondage n°2.

Mais par hypothèse W est fonction de ζ , donc $\mathcal{E}W(Z_1, \zeta)$ ne dépend pas des sondages (2), d'où l'obligation pour $k_2 - \mathcal{E}K_{21}$ de disparaître.

Donc $\mathcal{E}W$ est nécessairement de la forme $\mathcal{E}\phi(Z) - \phi(\zeta)$. cqfd.

Réciproque du Théorème I.

$$\begin{cases} \mathcal{E}W(Z_1, \zeta) = \mathcal{E}\phi(Z_1) - \phi(\zeta) \\ \mathcal{E}W(Z_2, Z_1) = \mathcal{E}\phi(Z_2) - \phi(Z_1) \end{cases}$$

$$\begin{aligned} \mathcal{E}W(Z_1, \zeta) + \mathcal{E}EW(Z_2, Z_1) &= \mathcal{E}\phi(Z_2) - \phi(\zeta) \\ &= \mathcal{E}W(Z_2, \zeta) \end{aligned}$$

A condition que $[\mathcal{E}\phi(Z) - \phi(\zeta)]$ ait un signe constant (qu'on rend au besoin positif en changeant ϕ de signe) il s'en suit bien que c'est une perte d'information. Mais la restriction qu'on vient de formuler a beaucoup d'importance.

Remarque 1.

$[\mathcal{E}\phi(Z) - \phi(\zeta)]$ est le biais quand on estime $\phi(\zeta)$ par $\phi(Z)$, [alors que Z estime ζ sans biais].

Remarque 2:

Par exemple de la relation $\mathcal{V}(\bar{X}) = \mathcal{E}(\bar{X}^2) - \bar{x}^2$, on déduit $\mathcal{E}(\bar{X}^2) - \bar{x}^2 \geq 0$ ou $\phi(Z) - \phi(\zeta) \geq 0$ avec $\zeta = \bar{x}$, $\phi(Z) = Z^2$, $Z = \bar{X}$.

Plus généralement : toute fonction monotone convexe ϕ satisfait à notre condition de signe; la distribution des points (avec $\mathcal{E}Z = \zeta$) $[Z, \phi(Z)]$ admet le point $[\zeta, \mathcal{E}\phi(Z)]$ comme barycentre, point situé dans la concavité de la courbe.

C'était le cas pour $\phi = \zeta^3$ (avec $\zeta > 0$) et $\phi = \zeta^4$, aux exemples du n°3.

Le cas des estimateurs biaisés sera étudié au paragraphe II ci-après.

Remarque 3.

Si W est une fonction dérivable en ζ , on peut poser $\omega = \partial W / \partial \zeta$ et dériver les deux membres de l'équation de Chasles par rapport à ζ en posant

$$\mathcal{E}W = \sum_i p^i W(z^i, \zeta)$$

et en traitant les p^i comme des constantes.

On a d'abord

$$\sum_k p_2^k \omega(\zeta, z_2^k) = \sum_h p_1^h \omega(\zeta, z_1^h)$$

d'où il suit que l'expression :

$$\sum_i p_1^i \omega(\zeta, z_1^i), \quad \text{avec Prob}(Z_j = z_j^i) = p_j^i$$

doit être indépendante du rang j que l'estimateur Z_j occupe dans la suite $\zeta Z_1 Z_2 \dots$ d'estimateurs relatifs à des échantillons de tailles décroissantes. Autrement dit :

$$\sum_i p_j^i \omega(\zeta, z_j^i) = f(\zeta)$$

Supposons $f(\zeta) \neq 0$:

Posons :
$$-\phi(\zeta) = \int f(\zeta) d\zeta$$

et intégrons la dérivée de l'équation de Chasles (avec les p^i constants). Il viendra

$$\sum_k p_2^k W(\zeta, z_2^k) = \int f(\zeta) d\zeta - a_2 = -\phi(\zeta) - a_2 \text{ (et analogue)}$$

avec
$$-\sum_h p_1^h \phi(z_1^h) = a_1 \text{ (et analogue)}$$

d'où :
$$-\phi(\zeta) + \sum_k p_2^k \phi(z_2^k) = -\phi(\zeta) + \sum_h p_1^h \phi(z_1^h) + \sum_h p_1^h [-\phi(z_1^h) + \sum_k p_2^k \phi(z_2^k)]$$

donc :
$$\mathcal{E} W(\zeta, Z) = \mathcal{E} \phi(Z) - \phi(\zeta)$$

On retrouve le théorème I. Il faut noter pourtant que $f(\zeta)$ n'est pas $\mathcal{E} \omega$, car on n'a pas le droit de remplacer

$$\frac{\partial}{\partial \zeta} \mathcal{E} W \quad \text{par} \quad \mathcal{E} \frac{\partial W}{\partial \zeta}.$$

En effet : si par exemple W est de la forme

$$A_0(\zeta) + A_1(\zeta)Z + A_2(\zeta)Z^2 + \dots$$

on aurait

$$\frac{\partial}{\partial \zeta} \mathcal{E} W = A_0'(\zeta) + A_1'(\zeta) \mathcal{E} Z + A_2'(\zeta) \mathcal{E} Z^2 + \dots + A_1(\zeta) \frac{\partial}{\partial \zeta} (\mathcal{E} Z) + A_2(\zeta) \frac{\partial}{\partial \zeta} (\mathcal{E} Z^2) + \dots$$

contre
$$\mathcal{E} \frac{\partial W}{\partial \zeta} = A_0'(\zeta) + A_1'(\zeta) \mathcal{E} Z + A_2'(\zeta) \mathcal{E} Z^2 + \dots$$

avec par exemple : $\mathcal{E} Z = \zeta$, donc $\frac{\partial}{\partial \zeta} (\mathcal{E} Z) = 1$, etc.

On a supposé $f(\zeta) \neq 0$.

Au contraire on a $\mathcal{E}\omega = 0$, c'est-à-dire $\mathcal{E}(\zeta - Z) = 0$ (absence de biais).

Finalement il est superflu de supposer W dérivable; et la première démonstration du théorème I est moins discutable que celle-ci.

4/ - THEOREME II a -

La perte d'information étant $\mathcal{E}W$, où W est une fonction de Z et ζ , il existe nécessairement entre estimateur Z et estimé ζ une relation de la forme

$$\mathcal{E}\Delta(Z, \zeta) = 0$$

En effet posons $W(Z, \zeta) = \Delta(Z, \zeta) + \phi(Z) - \phi(\zeta)$ avec le ϕ du théorème I ; il vient de la sorte

$$\mathcal{E}\Delta(Z, \zeta) = 0 \quad \text{cqfd.}$$

Cas particulier : $\mathcal{E}Z - \zeta = 0$ (estimation sans biais).

THEOREME II b -

S'il existe entre estimateur Z et estimé ζ une relation de la forme

$$\mathcal{E}\omega(Z, \zeta) = 0$$

on a nécessairement $\Delta(Z, \zeta) \equiv \psi(\zeta) \cdot \omega(Z, \zeta)$ (Δ étant celui défini au théorème II (a), sous la double condition :

- que la perte d'information soit suffisamment régulière;

- que la distribution sondée soit suffisamment générale.

(condition qui sera supposée remplie dans les numéros suivants sauf avis contraire).

En effet si par exemple ω et Δ sont deux fonctions de Z développables en série entière

$$\omega = \sum_0^{+\infty} a_i Z^i, \quad \Delta = \sum_0^{+\infty} b_i Z^i$$

les deux équations en ζ

$$\begin{cases} \mathcal{E}\omega \equiv \sum_0^{+\infty} a_i \mathcal{E}(Z^i) = 0, \\ \mathcal{E}\Delta \equiv \sum_0^{+\infty} b_i \mathcal{E}(Z^i) = 0, \end{cases}$$

déterminent le même ζ si (et seulement si) b_i/a_i est une constante, soit $\psi(\zeta)$ (en supposant donnés la distribution à sonder, le plan de sondage et l'estimateur et en excluant toute relation particulière entre $\mathcal{E}Z$, $\mathcal{E}Z^2$, $\mathcal{E}Z^3$, etc.).

Cas particulier : $\Phi(\zeta) = \zeta^2$;

$$\Delta = (Z - \zeta)^2 + \zeta^2 - Z^2 = 2\zeta^2 - 2\zeta Z = 2\zeta(\zeta - Z)$$

$$\omega = \zeta - Z ; \psi = 2\zeta$$

5/- Conséquence : Relations entre pertes conjuguées d'information

Lorsqu'on permute ζ et Z , la perte d'information n'est pas conservée si W n'est pas symétrique en Z et ζ ; on dira que ces deux pertes d'information sont conjuguées.

Désignant par une astérique les fonctions où Z et ζ sont permutés, on a (ψ désignant $\psi(\zeta)$ et ψ^* désignant $\psi(Z)$) :

$$W = \Phi(Z) - \Phi(\zeta) + \Delta$$

$$W^* = \Phi(\zeta) - \Phi(Z) + \Delta^*$$

d'où

$$W + W^* = \Delta + \Delta^* = \psi\omega + \psi^*\omega^*$$

$$W - W^* = 2[\Phi(Z) - \Phi(\zeta)] + (\Delta - \Delta^*)$$

Le cas où $W (= W^*)$ est symétrique est plus spécialement intéressant. On a alors :

$$\Phi(\zeta) - \Phi(Z) = (\Delta - \Delta^*)/2 = (\psi\omega - \psi^*\omega^*)/2$$

$$W = (\Delta + \Delta^*)/2 = (\psi\omega + \psi^*\omega^*)/2$$

6/- Résolution de l'équation de Chasles avec un estimateur sans biais

THEOREME III -

a) Si W est symétrique, on a

$$W = \lambda(Z - \zeta)^2, \quad \Phi = \lambda\zeta^2$$

En effet on a $\omega = \zeta - Z$; d'où : $\omega^* = -\omega$

$$\Phi(\zeta) - \Phi(Z) = (\psi\omega - \psi^*\omega^*)/2$$

ou

$$2\Phi(\zeta) - 2\Phi(Z) = (\zeta - Z)\psi(\zeta) - (Z - \zeta)\psi(Z)$$

c'est-à-dire :

$$[2\Phi(Z) - Z\psi(Z)] + [\zeta\psi(Z) - Z\psi(\zeta)] - [2\Phi(\zeta) - \zeta\psi(\zeta)] = 0$$

où chacun des crochets doit s'annuler séparément à cause de la grande généralité imposée aux distributions (donc à celle de Z); d'où :

$$2 \frac{\Phi(Z)}{Z} = \psi(Z); \quad \frac{\psi(Z)}{Z} = \frac{\psi(\zeta)}{\zeta}; \quad 2 \frac{\Phi(\zeta)}{\zeta} = \psi(\zeta)$$

d'où

$$\frac{\Phi(Z)}{Z^2} = \frac{\Phi(\zeta)}{\zeta^2}$$

imposé quels que soient Z et ζ , d'où

$$\Phi(Z) = \lambda Z^2, \quad \Phi'(\zeta) = \lambda \zeta^2, \quad \psi(\zeta) = 2\zeta$$

$$W = [2\lambda\zeta(\zeta - Z) + 2\lambda Z(Z - \zeta)]/2 = \lambda(Z - \zeta)^2 \quad \text{cqfd.}$$

b) Si W est asymétrique, W est nécessairement de la forme

$$W = \Phi(Z) - \Phi(\zeta) + \Phi'(\zeta) \cdot (\zeta - Z)$$

où Φ est arbitraire, à part la condition $\Phi'' > 0$.

En effet $\partial W / \partial Z = \Phi'(Z) - \psi(\zeta)$ doit être nul pour $Z = \zeta$; et $\partial^2 W / \partial Z^2 = \Phi''(Z)$ doit être positif, de façon à s'assurer que $W(Z, \zeta)$ autour de $W(\zeta, \zeta) = 0$, est positif et croît avec $|Z - \zeta|$ ⁽¹⁾. Il suit de là que

$$(\partial W / \partial Z)_0 = \Phi'(\zeta) - \psi(\zeta) = 0,$$

d'où le théorème énoncé.

7/ - Extension immédiate à un certain estimateur biaisé (moyennes à la Gini).

f étant une fonction arbitraire, substituant $f(\zeta)$ à ζ et $f(Z)$ à Z ,

(1) Nota : $\Phi'' > 0$ équivaut (lorsque Φ'' existe) à la convexité requise pour Φ ; voir réciproque du théorème I.

il vient les expressions suivantes de ω , de W et de Φ dans le cas où W est symétrique en Z et ζ , - de W dans le cas où W n'est pas symétrique (avec $\Phi'' > 0$)⁽¹⁾ :

$$\omega = f(\zeta) - f(Z)$$

$$a) \quad W = W^* = \lambda [f(Z) - f(\zeta)]^2, \quad \Phi(\zeta) = \lambda [f(\zeta)]^2$$

$$b) \quad W = \Phi[f(Z)] - \Phi[f(\zeta)] + \Phi'[f(\zeta)] \cdot [f(\zeta) - f(Z)]$$

II - LES PERTES D'INFORMATION DANS LE CAS DES ESTIMATIONS BIAISEES -

1/ - Introduction.

On vient d'étendre à l'estimation biaisée de ζ par Z , fruit de l'estimation sans biais de $f(\zeta)$ par $f(Z)$, le théorème sur la perte de l'estimation sans biais. Ceci ne présente qu'un intérêt pratique médiocre, car les estimateurs biaisés les plus employés ne sont pas de cette forme; mais l'intérêt est d'ordre théorique.

On va voir que les moyens mis en œuvre jusqu'ici sont alors en échec. Si la distribution de Z n'est absolument pas particularisée et si les fonctions considérées sont parfaitement régulières (et d'abord si ce sont bien des fonctions, non des fonctionnelles) le théorème qui suit montre qu'il n'y a rien au-delà de l'estimation sans biais de $f(\zeta)$ par $f(Z)$.

THEOREME 1 -

Pour qu'une fonction $W(Z, \zeta)$ existe, il est nécessaire que $\mathcal{E} \omega = 0$ soit de la forme

$$\psi(\zeta) [\mathcal{E} f(Z) - f(\zeta)] = 0$$

En effet

$$\omega(Z_1, \zeta) = \Omega \quad \text{s'écrit} \quad Z_1 = g(\Omega, \zeta)$$

$$\omega(Z_2, Z_1) = \Omega_1 \quad \text{s'écrit} \quad Z_2 = h(\Omega_1, Z_1) = h_1(\Omega_1, \Omega, \zeta)$$

d'où

$$\omega(Z_2, \zeta) = h_2(\Omega_1, \Omega, \zeta)$$

Or $\mathcal{E} \omega(Z_2, Z_1) = \mathcal{E} \Omega_1 = 0$ détermine Z_1 en fonction de la distribution de Z_2 ; de même $\mathcal{E} \omega(Z_1, \zeta) = \mathcal{E} \Omega = 0$ détermine ζ . Ces deux relations entraînent donc

$$\mathcal{E} \omega(Z_2, \zeta) = 0$$

(1) Voir note de la page précédente.

Ceci n'est possible que si l'on a :

$$\omega(Z_2, \zeta) = \lambda_1 \Omega_1 + \lambda \Omega$$

(en supposant h_2 suffisamment régulière et les distributions de Ω_1 et Ω assez générales); soit

$$\omega(Z_2, \zeta) = \lambda_1 \omega(Z_2, Z_1) + \lambda \omega(Z_1, \zeta)$$

qui n'est possible qu'avec $\lambda = \lambda_1 = 1$ (faire $\zeta = Z_1$ ou $Z_1 = Z_2$); ce qui implique

$$\omega(Z_2, Z_1) = f(Z_2) - f(Z_1) \quad \text{cqfd.}$$

2/ - Cas où la fonction $W(Z)$ ne serait pas fonction de ζ .

Puisqu'on renonce à se limiter aux W symétriques, il est naturel de supposer que W est seulement fonction de Z , avec $W(\zeta) = 0$. D'ailleurs le texte de Basu cité à la fin du chapitre III, § I nous y invite. On constate alors ce qui suit.

THEOREME 2.

Si W n'est pas fonction de ζ , et est fonction continue et dérivable de Z , l'équation de Chasles est vérifiée, à la condition (nécessaire et suffisante) que $W(z)$ soit une constante.

En effet la relation de Chasles prend alors la forme :

$$\mathcal{E} W_0(Z_1) = \mathcal{E} W_0(Z_2) - \mathcal{E} W_h(Z_2)$$

où l'indice 0 indique que la distribution de base est la population (des x_i); alors que l'indice (h) désigne le grand échantillon fournissant l'estimation $Z_1 = z_1^h$ et sur lequel est calculée la perte d'information $W_h(z_2^k)$ pour chaque petit échantillon (k).

$\mathcal{E} W_0(Z_1)$ est fonction des z_1^h et indépendante des z_2^k .

Mais si W_0 n'est pas fonction de ζ , W_h n'est pas fonction de z_1^h ; et il s'en suit que le second membre a sa dérivée nulle par rapport à z_1^h ; donc aussi le premier membre.

Par conséquent $W_0(z_1^h)$ resterait constant, au moins lorsqu'on donne à z_1^h de petites variations.

Si nous excluons le cas où $W(Z)$ serait discontinue, on a bien :

$$W_0(z_1^h) = \lambda^h, \quad \text{donc } \mathcal{E} W_0(z_1^h) = \mathcal{E}(\lambda^h) = \Lambda_1$$

$$\mathcal{E} W_0(z_2^k) = \Lambda_2$$

$$\mathcal{E} E W_h(z_2^k) = \Lambda_2 - \Lambda_1$$

Remarque.

Supposer $W(z)$ continue, c'est écarter le cas où $W(z)$, tout en étant "globalement" variable, resterait constante dans des intervalles, autrement dit augmenterait par saut quand z s'éloigne de ζ . En effet, on peut considérer qu'un tel indicateur de la perte d'information serait défectueux du point de vue conceptuel (tout en acceptant en revanche que la courbe (convexe) représentant $W(z)$ présente des points anguleux - par exemple soit polygonale).

3/ - Le cas exceptionnel d'estimation biaisée.

On rencontre très facilement le cas suivant, qui correspond à une distribution de Z tout à fait particulière :

THEOREME -

Si l'erreur d'échantillonnage a un signe constant, tant pour $Z_2 - Z_1$ que pour $Z_1 - \zeta$, le module du biais est perte d'information, ainsi que tout $|g(Z) - g(\zeta)|$, où g désigne une fonction monotone.

Réciproquement, si $\mathcal{E} Z - \zeta$ a un signe constant, il est nécessaire que l'erreur ait un signe constant pour qu'on puisse écrire : $W = Z - \zeta$.

En d'autres termes, Z_1 ne prend que des valeurs z_1^h disons inférieures ou égales à ζ ; $Z_2(z_1^h)$ ne prend que des valeurs inférieures ou égales à z_1^h ; etc.

Bref l'équation de Chasles est vérifiée, pour la seule raison qu'on a (terme à terme) :

$$W(z_2^k, \zeta) = W(z_1^h, \zeta) + W(z_2^k, z_1^h)$$

$$\text{en posant} \quad \pm W(z, \zeta) = \zeta - z$$

$$\text{voire} \quad \pm W(z, \zeta) = g(\zeta) - g(z).$$

Bref la perte d'information est anti symétrique. En voici deux applications très simples.

On s'est limité aux sondages exhaustifs.

4/ - Estimation de la plus grande valeur des x_i .

Soit $a > b > c \dots$ les valeurs des x_i . Soit $f(x)$ une fonction croissante (sur tout l'intervalle des x_i) avec, par définition :

$$W(x_i, a) = (ax_i) = f(a) - f(x_i)$$

par exemple

$$(ax_i) = a - x_i$$

Application.

Pour $n = v - 1$, il y a une chance sur v que l'échantillon ne renferme pas (a), mais alors il renferme forcément (b) et la perte d'information est égale à

$$\mathcal{E} W = \left[\frac{1}{v} (ab) \right]$$

Pour $n = v - 2$: parmi $C_v^{v-2} = C_v^2$ combinaisons, dont $(v - 1)$ renfermant la lettre a, les combinaisons ne renfermant pas (a) mais (b) sont au nombre de $(v - 2)$; et une combinaison ne renferme ni (a) ni (b) mais (c). Ainsi la perte d'information est :

$$\mathcal{E} W = [(v - 1) (ab) + (ac)] / C_v^2$$

Pour n quelconque : sur C_v^n combinaisons : C_{v-1}^n renfermant a, C_{v-2}^n renfermant b (sans a); C_{v-3}^n renfermant c (sans a ni b); etc.

D'où la perte d'information :

$$\left[C_{v-2}^n (ab) + C_{v-3}^n (ac) + \dots \right] / C_{v-1}^n$$

Remarques.

1) La théorie précédente se transpose immédiatement au cas de l'estimation de la plus petite des valeurs x_i .

2) Dans les deux cas, l'estimateur Z de ζ est toujours biaisé, - par défaut s'il s'agit de la plus grande valeur, - par excès dans l'autre alternative; l'écart ζZ est toujours de même signe.

3) A mesure que l'échantillon devient plus petit, Z ne peut que s'éloigner davantage de ζ ; autrement dit ζZ_n et $Z_n Z_{n-1}$ sont de même signe.

Si la plus grande valeur connue pour un certain échantillon(n) est (b), la plus grande valeur pour les n échantillons d'effectif (n - 1) qu'il engendre sera b, ou c, ou d ... mais en aucun cas ne peut devenir a.

5/ - Estimation de l'étendue (ou amplitude) ("range").

L'étendue (r), écart entre le plus grand et le plus petit des x_i , ne peut être estimée que si l'échantillon comprend au moins deux boules, $n \geq 2$, en quoi elle diffère de la plus grande (a) des valeurs de x_i , qui demande seulement qu'on ait : $n \geq 1$.

Cela mis à part, r et a sont des propriétés tout à fait analogues : lorsque l'effectif n de l'échantillon diminue, depuis v jusqu'à 2, l'écart entre le plus grand et le plus petit des x_i - échantillons est au plus égal à r et ne peut jamais augmenter. Il fournit donc de r un estimateur biaisé par défaut.

La perte d'information correspondante peut être définie de façon comparable.

Soit $a > b \dots > y > z$, les valeurs de x_i .

Pour $n = v - 1$, il y a $v - 2$ chances sur v que $(a'z') = (az)$.

$$W = \frac{1}{v} [(ab) + (yz)]$$

Pour $n = v - 2$:

$$W = \frac{1}{C_v^n} \{ 0 \cdot C_{v-2}^2 + [(ab) + (yz)] C_{v-2}^1 + [(ac) - (xz)] \}$$

Pour n quelconque :

$$W = \frac{1}{C_v^n} \{ 0 \cdot C_{v-2}^n + [(ab) + (yz)] [C_{v-3}^{n-1} + C_{v-4}^{n-2}] + [(ac) + (xz)] [C_{v-3}^{n-2} + \dots + C_{v-6}^{n-4}] + \dots \}$$

6/ - Retour au cas général de l'estimation biaisée.

Alors que ni $VZ = \mathcal{E}Z^2 - \zeta^2 + [\zeta^2 - (\mathcal{E}Z)^2]$,

$$\text{ni } \mathcal{E}(Z - \zeta)^2 = \mathcal{E}Z^2 - \zeta^2 + 2\zeta(\zeta - \mathcal{E}Z)$$

ne sont plus des pertes d'information, il en existe d'évidentes, en ce sens que la relation de récurrence est manifestement vérifiée : telles sont

$$\mathcal{E}Z^2 - \zeta^2 \quad \text{ou} \quad \mathcal{E}\Phi(Z) - \Phi(\zeta)$$

où Φ est une fonction arbitraire, sous réserve (a) qu'on ait bien :

$$\Phi(\zeta) < \mathcal{E}\Phi(Z_1) < \mathcal{E}\Phi(Z_2) \dots$$

avec les notations du théorème I (§I), et (b) que Z_2 se réduise bien à Z_1 , lorsque l'échantillon de π_2 recouvre celui de π_1 (les plans de sondage π_1 et π_2 appartenant à une suite de plans à effectifs décroissants pris sur la famille $\mathcal{F}$).

La restriction (a) n'est plus une banale concavité puisque ζ n'est plus la moyenne de Z_1 . La restriction (b) est une condition d'estimation "cohérente" au sens des sondages (laquelle ne comporte ni existence de tirages infiniment nombreux ni convergence en probabilité). L'estimateur Z (biaisé) sera donc supposé à présent "uniformément cohérent", c'est-à-dire cohérent tout au long des suites de plans telles que

$$[U \pi_1 \pi_2 \dots]$$

On s'intéressera ici plus spécialement aux estimateurs isomorphes du paramètre à estimer. La notion d'estimation uniformément cohérente est un peu plus large que l'estimation isomorphe (l'expression de ζ à laquelle se réduit Z lorsque l'échantillon de π recouvre l'univers U pouvant être finalement plus simple que celle de Z). Toute fonction arbitraire $g(Z)$ est d'ailleurs estimateur isomorphe de $g(\zeta)$, ce qui explique l'arbitraire laissé à la fonction Φ .

La première difficulté est celle de l'affectation de telle ou telle perte d'information à tel ou tel estimateur. Pourquoi $(\mathcal{E}Z^2 - \zeta^2)$ serait-elle nécessairement affectée à l'estimation de ζ par Z . Quand on estime le ratio $\zeta = a/b$ par $Z = A/B$ avec $\mathcal{E}A = a$; $\mathcal{E}B = b$, on sait que les termes en (n^{-1}) de $\mathcal{E}(Z - \zeta)^2$ et du biais sont :

$$\mathcal{E}(Z - \zeta)^2 \# (\gamma^2 - 2\rho\gamma\gamma' + \gamma'^2) \cdot n^{-1}$$

$$\mathcal{E}Z - \zeta \# (\gamma'^2 - \rho\gamma\gamma') \cdot n^{-1}$$

Notons que, de toute façon, $V(Z)$ et $\mathcal{E}(Z - \zeta)^2$ sont équivalents :

$$\mathcal{E}(Z - \zeta)^2 = V(Z) + (\mathcal{E}Z - \zeta)^2$$

En revanche $\mathcal{E}Z^2 - \zeta^2 = \mathcal{E}(Z - \zeta)^2 + 2\zeta(\mathcal{E}Z - \zeta)$

de sorte que :

$$\mathcal{E}Z^2 - \zeta^2 \# [\gamma^2 - 2\rho\gamma\gamma'(1 + \zeta) + \gamma'^2(1 + 2\zeta)] \cdot n^{-1}$$

Nous serons amené plus loin à définir une perte d'information

pour Z qui est en désaccord avec ce résultat; autrement dit nous n'attacherons ($\mathcal{E}Z^2 - \zeta^2$) à Z que lorsque le biais est nul (ou d'ordre supérieur au premier). Ceci montre qu'une certaine prudence est nécessaire pour rester cohérent.

7/ - Théorie de l'estimation isomorphe.

Désignons par $W[Z, \zeta]$ la fonctionnelle qu'on entend substituer à la fonction $W(Z, \zeta)$.

Posons d'une manière générale :

$$\mathcal{E}W[Z, \zeta] = \mathcal{E}\Phi[Z] - \Phi[\zeta]$$

$$W[Z, \zeta] = \Phi[Z] - \Phi[\zeta] + \Delta[Z, \zeta]$$

(les crochets désignant des fonctionnelles). On aura nécessairement :

$$\mathcal{E}\Delta[Z, \zeta] = 0, \quad W[\zeta, \zeta] = 0, \quad W[Z, \zeta] > 0$$

plus une condition de convexité pour W .

1er problème.

Traduire par une condition de la forme

$$\mathcal{E}\Delta[Z, \zeta] = 0$$

la relation entre estimateur Z et estimé ζ .

S'il s'agit de $\zeta = a/b$ et $Z = A/B$, avec

$$\mathcal{E}A - a = \mathcal{E}B - b = 0,$$

en raisonnant comme pour établir le théorème 1, on montrera qu'on a nécessairement (si Δ est fonction très régulière de A et B) :

$$\Delta[Z, \zeta] = \psi_1[\zeta] (a - A) + \psi_2[\zeta] (b - B)$$

D'une façon générale on va supposer que $\zeta = f(a, b, c, \dots)$ et $Z = f(A, B, C, \dots)$ avec

$$\mathcal{E}A = a; \quad \mathcal{E}B = b; \quad \mathcal{E}C = c \dots$$

- l'isomorphisme de Z et ζ est traduit par la présence d'une fonction commune f ;

- les $A B C \dots$ sont calculés sur l'échantillon et les $a b c \dots$ sur l'univers;

- A est isomorphe de a, B de b, C de c, etc.

Alors on a nécessairement (pourvu que Δ soit une fonctionnelle très régulière) :

$$\Delta[Z, \zeta] = \psi_1[\zeta] \cdot (a - A) + \psi_2[\zeta] \cdot (b - B) + \psi_3[\zeta] \cdot (c - C) + \dots$$

d'où

$$W[Z, \zeta] = \phi[Z] - \phi[\zeta] + \psi_1[\zeta] (a - A) + \psi_2[\zeta] (b - B) + \psi_3[\zeta] (c - C) + \dots$$

2ème problème: Cas où $W[Z, \zeta]$ est symétrique en Z et ζ .

Le calcul du théorème III se transpose : $W = W^*$; posons $\psi[\zeta] = \psi$, $\psi[Z] = \psi^*$.

$$2\{\phi[\zeta] - \phi[Z]\} \equiv (a - A)(\psi_1 + \psi_1^*) + (b - B)(\psi_2 + \psi_2^*) + (c - C)(\psi_3 + \psi_3^*) \dots$$

$$\begin{cases} 2\phi[\zeta] \equiv a\psi_1 + b\psi_2 + c\psi_3 \dots \\ 0 \equiv (a\psi_1^* - A\psi_1) + (b\psi_2^* - B\psi_2) + (c\psi_3^* - C\psi_3) \dots \\ 2\phi[Z] \equiv A\psi_1^* + B\psi_2^* + C\psi_3^* \dots \end{cases}$$

$$\frac{\psi_1}{a} = \frac{\psi_1^*}{A} = \alpha; \quad \frac{\psi_2}{b} = \frac{\psi_2^*}{B} = \beta; \quad \frac{\psi_3}{c} = \frac{\psi_3^*}{C} = \gamma; \dots$$

$$\text{d'où} \quad 2\phi[Z] \equiv \alpha A^2 + \beta B^2 + \gamma C^2 \dots$$

$$2W[Z, \zeta] = \alpha(A - a)^2 + \beta(B - b)^2 + \gamma(C - c)^2 + \dots$$

Cette solution n'est pas contestable; mais il se trouve que ce n'est pas celle qui est employée dans la routine des sondages.

3ème problème.

Traduire par une relation entre ϕ et les ψ_i la convexité de W.
On a :

$$\begin{aligned} \frac{\partial W}{\partial A} &= \frac{\partial \phi}{\partial A} + \frac{\partial \Delta}{\partial A} \\ &\equiv \frac{\partial \phi}{\partial A} - \psi_1[\zeta] \end{aligned}$$

en supposant que $W[Z]$ est en réalité une fonction dérivable $W(A, B, C \dots)$.

On va écrire que, pour $\zeta = Z$ (c'est-à-dire $a = A$, $b = B$, $c = C$), on a :

$$\partial W / \partial A = 0, \quad \partial W / \partial B = 0, \quad \partial W / \partial C = 0 \dots$$

$$\text{d'où : } \psi_1 = \partial \Phi [\zeta] / \partial a, \quad \psi_2 = \partial \Phi [\zeta] / \partial b, \quad \psi_3 = \partial \Phi [\zeta] / \partial c \dots$$

$$\text{et } W = \Phi [Z] - \Phi [\zeta] + \frac{\partial \Phi [\zeta]}{\partial a} (a - A) + \frac{\partial \Phi [\zeta]}{\partial b} (b - B) + \frac{\partial \Phi [\zeta]}{\partial c} (c - C) \dots$$

relation transposée de celle du théorème III. De même on doit écrire

$$\frac{\partial^2 \Phi}{\partial a^2} \geq 0, \quad \frac{\partial^2 \Phi}{\partial b^2} \geq 0, \quad \frac{\partial^2 \Phi}{\partial c^2} \geq 0 \dots$$

avec

$$\frac{\partial^2 \Phi}{\partial a^2} \quad \frac{\partial^2 \Phi}{\partial b^2} - \left(\frac{\partial^2 \Phi}{\partial a \partial b} \right)^2 \geq 0 \text{ etc.}$$

En particulier en supposant symétrique en ζ , Z la fonctionnelle $W [Z, \zeta]$ on retrouve le résultat précédent.

Autre solution particulière.

$$\text{Posons } \psi_1 = \psi \alpha, \quad \psi_2 = \psi \beta, \quad \psi_3 = \psi \gamma \dots$$

Il vient :

$$2(\Phi [\zeta] - \Phi [Z]) = (\psi + \psi^*) \{ \alpha (a - A) + \beta (b - B) + \gamma (c - C) \dots \}$$

en supposant que $\alpha \beta \gamma \dots$ ne sont pas transposés car constants; on posera :

$$\alpha (a - A) + \beta (b - B) + \gamma (c - C) \dots = \omega [Z, \zeta]$$

avec

$$\omega^* = -\omega$$

et le calcul du théorème III donne cette fois

$$W = \Phi [Z] - \Phi [\zeta] + 2 \sqrt{\Phi} [\zeta] \{ \sqrt{\Phi} [\zeta] - \sqrt{\Phi} [Z] \}$$

avec

$$\sqrt{\Phi} [Z] = \alpha A + \beta B + \gamma C = \psi [Z]$$

$$\sqrt{\Phi} [\zeta] = \alpha a + \beta b + \gamma c = \psi [\zeta]$$

La perte d'information est :

$$\mathcal{E} W = \mathcal{E} [\alpha(A - a) + \beta(B - b) + \gamma(C - c)]^2$$

Choix de $\alpha\beta\gamma$.

Jusqu'ici le choix de $\alpha\beta\gamma$ était resté arbitraire. Les conditions de convexité étaient bien vérifiées ainsi :

$$\partial\Phi/\partial a = 2\alpha(\alpha a + \beta b + \gamma c)$$

d'où

$$\frac{\partial^2 \Phi}{\partial a^2} = 2\alpha^2 > 0 \quad \left(\frac{\partial^2 \Phi}{\partial a \partial b} \right)^2 = 4\alpha^2 \beta^2 \equiv \frac{\partial^2 \Phi}{\partial a^2} \cdot \frac{\partial^2 \Phi}{\partial b^2}$$

On va particulariser $\alpha\beta\gamma$ de façon à rejoindre la définition du §1 lorsque $(A - a), (B - b), (C - c)$ sont infiniment petits. Pour cela on différencie $\zeta = f(a, b, c \dots)$; il vient :

$$d\zeta = f'a \cdot da + f'b \cdot db + f'c \cdot dc \dots$$

et on choisit α, β, γ proportionnels à $f'a, f'b, f'c$. Ainsi on substitue à

$$\psi(\zeta) = 2\lambda \cdot \zeta \quad (\text{théorème III})$$

la relation

$$d\psi[\zeta] = 2\lambda \cdot d\zeta$$

D'où la perte d'information (particulière, non symétrique) :

$$\mathcal{E} W = \lambda^2 \mathcal{E} [f'a \cdot (A - a) + f'b \cdot (B - b) + f'c \cdot (C - c)]^2$$

8/ - Application et conséquences.

1) Estimation du ratio.

$$\zeta = a/b \quad \text{d'où} \quad \log \zeta = \log a - \log b. \quad \frac{d\zeta}{\zeta} = \frac{da}{a} - \frac{db}{b}$$

$$\mathcal{E} W = \lambda^2 \zeta^2 \cdot \mathcal{E} \left(\frac{A - a}{b} - \frac{B - b}{b} \right)^2 \equiv \lambda^2 \zeta^2 \mathcal{E} \left(\frac{A}{a} - \frac{B}{b} \right)^2$$

L'expression usuelle des statisticiens est $\mathcal{E} \left(\frac{A}{a} - \frac{B}{b} \right)^2$, premier terme du développement de

$$v \left(\frac{A}{B} \right)$$

2) Estimation du coefficient de corrélation.

$$\rho = \text{Cov}(x y) / \sigma(x) \cdot \sigma(y)$$

On posera

$$a = N\sigma^2(x)/N - 1; \quad b = N\sigma^2(y)/N - 1; \quad c = N \text{Cov}(x, y)/N - 1$$

dont les estimateurs isomorphes A , B et C sont sans biais.

Soit $\rho^2 = \zeta$. Ainsi à $[a, b, c] - c^2 = 0$ correspond bien $[A, B, C] - C^2 = 0$.

$$d\zeta/\zeta = 2(dc/c) - da/a - (db/b)$$

$$\text{d'où} \quad \mathcal{E}W = \lambda^2 \zeta^2 \mathcal{E} \left(2 \frac{C}{c} - \frac{A}{a} - \frac{B}{b} \right)^2$$

à un facteur près (indépendant de Z).

Là encore il s'agit d'une première approximation de la variance.

Notation.

Au lieu de $\mathcal{E}W = \mathcal{E}(Z - \zeta)^2$ on peut donc écrire

$$\mathcal{E}W = \mathcal{E}(\delta Z - \delta \zeta)^2$$

Parmi toutes les pertes d'information (dépendant d'une fonctionnelle arbitraire ϕ) on a donc choisi plus particulièrement celle qui (quadratique en A, B, C) est asymptotique à la variance (lorsque Z est très voisin de ζ).

Remarque (complétant le § 6 ci-dessus).

C'est pour une raison de convexité que les estimateurs biaisés Z n'admettent pas $Z^2 - \zeta^2$ comme perte d'information.

Il suffit de le vérifier dans le cas du ratio :

$$Z = A/B, \quad \zeta = a/b, \quad \mathcal{E}A - a = \mathcal{E}B - b = 0$$

$$\text{Posons} \quad \phi(\zeta) = \zeta^2, \quad \mathcal{E}W = \mathcal{E}Z^2 - \zeta^2$$

$$W = \left(\frac{A}{B} \right)^2 - \left(\frac{a}{b} \right)^2 + \Delta, \quad \mathcal{E}\Delta = 0$$

$$\text{d'où} \quad \Delta = \lambda(a - A) + \mu(b - B)$$

Imposons à $W[Z] = W(A, B)$ d'être représentée par une surface tangente (au point $A = a, B = b$) au plan $\Delta = 0$; il vient

$$\frac{\partial W}{\partial A} = 2 \frac{A}{B^2} - \lambda, \quad \frac{\partial W}{\partial B} = - \frac{2 A^2}{B^3} - \mu$$

$$\text{d'où} \quad \lambda = 2 a/b^2, \quad \mu = - 2 a^2/b^3$$

$$\Delta \equiv \frac{2 a^2}{b^2} \left(\frac{B}{b} - \frac{A}{a} \right)$$

Montrons que la surface représentant $W(A, B)$ n'est pas convexe.
On a :

$$\frac{\partial^2 W}{\partial A^2} \cdot \frac{\partial^2 W}{\partial B^2} - \left(\frac{\partial^2 W}{\partial A \partial B} \right)^2 = \frac{2}{B^2} \frac{6 A^2}{B^4} - \left(\frac{4 A}{B^3} \right)^2 = - \frac{4 A^2}{B^6} < 0, \text{ cqfd}$$

D'ailleurs l'expression exacte de $\mathcal{E} Z^2$ est souvent fort compliquée; et au point de vue des calculs pratiques on gagne à renoncer à $\mathcal{E} Z^2 - \zeta^2$.

9/ - Estimation du maximum de vraisemblance.

a) On sait que, pour une variable Z dont la loi de distribution admet la densité

$$g(z, \zeta)$$

ζ étant le paramètre à estimer, la quantité d'information de Fisher (pour n tirages sans remise) est nH , avec

$$Y = \frac{\partial}{\partial \zeta} \log g(Z, \zeta), \quad H = \mathcal{E}(Y^2)$$

b) En fait (H/n) est une perte d'information; car c'est la variance d'une moyenne arithmétique $\bar{Y}$ d'espérance mathématique nulle. En effet, posons

$$y_i = \frac{\partial}{\partial \zeta} \log g(z_i, \zeta), \quad \bar{Y} = S y_i / n.$$

En dérivant (par rapport à ζ) sous le signe $\int$

$$\int_{-\infty}^{+\infty} g(z, \zeta) dz = 1$$

(à supposer qu'on ait le droit de le faire) on trouve

$$\mathcal{E} Y = 0; \quad \text{d'où} \quad \mathcal{E} \bar{Y} = 0. \quad \text{cqfd}$$

c) On peut considérer que l'estimation Z^0 du maximum de vraisemblance est un estimateur isomorphe de ζ . Plus précisément ζ et Z^0 sont racines de deux équations isomorphes, à savoir :

$$\psi(\zeta) = \mathcal{E} y(\zeta) = 0;$$

$$\Psi(Z) = S y_i(Z^0)/n = 0;$$

et, si par exemple leurs premiers membres sont développables en série entière :

$$\psi(t) = \mathcal{E} y(t) = a_0 + a_1 t + a_2 t^2 + \dots$$

$$\Psi(t) = \frac{1}{n} S y_i(t) = A_0 + A_1 t + A_2 t^2 + \dots$$

on a (quel que soit t) $\mathcal{E} \Psi = \psi$, et par conséquent $\mathcal{E} A_j = a_j$.

On se trouve donc dans un cas à peine plus général qu'aux $n^o 1$ et 2 : à $A B C \dots$ est substituée l'infinité des A_j .

d) Ainsi la perte d'information serait de la forme $\mathcal{E} W [Z^0]$, avec

$$W [Z^0] = \Phi [Z^0] - \Phi [\zeta] + \sum \lambda_j (a_j - A_j)$$

e) A ζ et g donnés correspond Z^0 ; donc ζ est fonctionnelle de Z^0 ; donc en faisant $t = \zeta$, $\Psi(\zeta)$ est une fonctionnelle de Z^0 qu'on peut appeler $\sqrt{\Phi} [Z^0]$.

Alors $\sqrt{\Phi} [\zeta]$ n'est autre que $\psi(\zeta)$, c'est-à-dire 0. D'où

$$\mathcal{E} \Phi [Z^0] - \Phi [\zeta] = \mathcal{E} \bar{Y}^2 - 0 = H/n$$

L'information de Fisher est donc le produit par n^2 d'une perte d'information relative à l'estimation (biaisée, isomorphe) du maximum de vraisemblance.

Nota - Cette variance ne pouvait représenter la perte d'information relative à l'estimation de $\mathcal{E} Y$ puisque $\mathcal{E} Y = 0$.

10/ - Généralisation : Variances de statistiques hors estimation.

Revenons sur le cas de n tirages équiprobables dans une urne unique de taille v . Soit Y_n une statistique quelconque (calculée sur l'échantillon);

$$\mathcal{E} Y_n = y(n, v)$$

dépend en général de n ; et on ne peut donc dire de Y_n que c'est un estimateur (sans biais, d'un quelconque η). En revanche $\forall Y_n$ est perte d'information du moment que la relation de récurrence est satisfaite, c'est-à-dire, avec $n > n'$, dès que

$$\mathcal{E}\{[Y_n - y(n, v)] E_n[Y_n - Y_n]\} = 0$$

a) Supposer $E_n[Y_n] = Y_n$ (uniformément) reviendrait à faire dépendre Y_n et y des $n' < n$, ce qui est à rejeter. En revanche rien n'empêche d'admettre que la différence

$$E_n Y_n - Y_n$$

est constante (ne dépend pas de Y_n) pour n' donné; ceci implique pour-tant que :

$$E_n Y_n - Y_n = y(n', v) - y(n, v),$$

ou
$$E_n\{Y_n - y(n', v)\} = Y_n - y(n, v).$$

Conclusion.

Il y a lieu de supposer la variable

$$Y_n - y(n, v)$$

uniformément sans biais, pour que la perte d'information $\forall Y_n$ existe (sans être encore attachée à quelque η).

Alors se pose le même problème qu'au n°9 : Peut-on rattacher

$$\lambda \{ \forall Y_n - y^z(n, v) \}$$

à l'estimation biaisée d'un certain ζ par un certain Z ?

Refaisant à l'envers le calcul du n°9, il vient

$$\psi(\zeta) = 0, \quad \Psi(\zeta) = Y_n - y(n, v)$$

Si Y_n dépend d'un paramètre ζ à estimer, tel que

$$\mathcal{E} Y_n(\zeta) - y(n, v, \zeta) = 0$$

l'estimateur "isomorphe" (au sens du n°9) est la racine de l'équation

$$Y_n(Z) - y(n, v, Z) = 0$$

b) Supposons plus simplement que $Y_n = S y_i / n$;

$$\mathcal{E} y_i = y = 0$$

où y_i dépend d'un paramètre ζ ; $\forall Y_n - \bar{y}^z$ est alors une perte d'information. Il vient

$$\Psi(\zeta) = \frac{\sum y_i}{n}$$

Posons :

$$\Psi(Z) = \frac{\sum y_i(Z)}{n} = 0$$

Il y a isomorphisme entre les équations $\mathcal{E}y = 0$ et $\sum y_i/n = 0$, et (au même sens qu'au n°9) entre ζ et son estimation Z .

Remarques.

1) Dans quels cas la variable $Y_n - y(n, v)$ est-elle uniformément sans biais ?

Considérons par exemple $Y_n = S^2$,

$$y(n, v) = \frac{(n-1)v}{(v-1)n} \sigma^2, \quad (Y_{n'} = S'^2).$$

On n'est pas dans le cas recherché, vu que

$$E \left[S'^2 - \frac{(n'-1)v}{(v-1)n'} \sigma^2 \right] = \frac{(n'-1)n}{(n-1)n'} \left[S^2 - \frac{(n-1)v}{(v-1)n} \sigma^2 \right]$$

La seule possibilité est d'avoir $E[Y_{n'}] = Y_n + k(n')$, k étant constant (quand n' est donné) c'est-à-dire indépendant de Y_n ; il en est ainsi quand Y_n est de la forme

$$\bar{X}_n + \sum_{v=1}^n \delta(v, i). \quad (\text{condition suffisante})$$

2) On peut (sous des conditions assez larges) écrire un développement tel que :

$$y(n, v) = \zeta + a_1 \left(\frac{1}{n} - \frac{1}{v} \right) + a_2 \left(\frac{1}{n} - \frac{1}{v} \right)^2 + \dots$$

et on n'a pas en général $a_2 = \dots = 0$. En outre un cas encore plus intéressant serait celui où

$$y(n, v) = \zeta + a_0 \left(\frac{1}{n} - \frac{1}{v} \right)^{1/2} + \dots$$

11/ - Quelques considérations générales sur l'estimation.

a) Il n'est pas question d'exiger de tout estimateur l'absence de biais; mais on ne peut accepter pour estimer ζ n'importe quelle statistique Z .

En pratique on a rencontré trois cas :

- isomorphisme et absence de biais (exemple : $nS^2/n - 1$ pour $v\sigma^2/v - 1$)
- isomorphisme avec biais (exemple : S^2 pour σ^2)
- absence de biais sans isomorphisme (exemple : $\frac{n(v - 1)S^2}{(n - 1)v}$ pour σ^2)

b) Les exemples ci-dessus relèvent du cas où, $\eta(v)$ étant isomorphe de l'aléatoire Y , on a

$$\mathcal{E}Y \equiv y(v, n) = \eta(v) \cdot a(v, n)$$

Un cas également simple est celui où on aurait :

$$\mathcal{E}Y = \eta(v) + b(v, n);$$

on poserait $\zeta = \eta(v)$, $Z = Y - b(v, n)$. Le cas mixte serait $\mathcal{E}Y = a\eta + b$, d'où $\zeta = \eta(v)$ pour $Z = [Y - b(v, n)]/a(v, n)$.

c) Mais en général il n'est pas question d'éliminer totalement le biais. Par exemple, que σ soit estimé au moyen de S , ou de $S\sqrt{n(v - 1)/v(n - 1)}$, il y a biais.

De ce qui précède, il résulte qu'on mesure la perte d'information relative à $\frac{nS^2}{n - 1}$ estimateur de $\frac{v\sigma^2}{v - 1}$, par $\mathcal{V}\left(\frac{nS^2}{n - 1}\right)$ relative à S^2 estimateur de $\frac{(n - 1)v}{(v - 1)n}\sigma^2$, par $\mathcal{V}S^2$; celle relative à S^2 , estimateur isomorphe biaisé de σ^2 , ne sera pas

$$\begin{aligned}\mathcal{E}(S^2 - \sigma^2)^2 &\equiv \mathcal{V}S^2 + \left[\frac{(n - 1)v}{(v - 1)n} \sigma^2 - \sigma^2 \right]^2 \\ &\equiv \mathcal{V}S^2 + \left[\frac{v\sigma^2}{v - 1} \right]^2 \left(\frac{1}{n} - \frac{1}{v} \right)^2\end{aligned}$$

$$\text{mais } \lambda \mathcal{V}\left(\frac{nS^2}{n - 1}\right)$$

La perte d'information n'est jamais définie qu'à un facteur constant près; et il y a lieu de considérer que $nS^2/n - 1$, S^2 , et aussi S , ont pour perte d'information symétrique $\lambda \mathcal{E}\left(\frac{nS^2}{n - 1} - \frac{v\sigma^2}{v - 1}\right)^2$. Le choix de λ demande une convention supplémentaire (voir Chapitre VII);

d) On a vu (Ch. III, § I, C) que Basu trouvait moins de risque à estimer $\bar{x}$ par $c \bar{X}$ ($c < 1$) que par $\bar{X}$. Au point de vue "perte d'information", il n'y aurait pas de différence entre $\bar{X}$ et $c \bar{X}$ (ou plutôt $c \bar{X}$ n'informe pas sur $\bar{x}$).

On verra au chapitre VII que certains estimateurs permettent de réduire les pertes d'information (théorie des informations supplémentaires); ce ne serait pas le cas de $c \bar{X}$.

On ne saurait pourtant se passer d'un critère pour le choix de l'estimateur ($c = 1$, par exemple); critère comprenant comme cas particulier l'absence de biais.

e) La coïncidence complète des formes analytiques de Z et ζ (source du mot "isomorphisme") n'est pas nécessaire en soi; elle assure a fortiori la "cohérence" quand l'échantillon grandit jusqu'à recouvrir toute la population; mais la cohérence (consistency) elle-même lorsque l'échantillon est petit, n'est pas toujours exigée (et peut empêcher l'élimination d'un biais, auquel cas on renonce à la cohérence).

f) En revanche la relation de récurrence de l'axiomatique de l'information adoptée ici suppose (si Z et Z' sont estimateurs de ζ dans une suite de plans de sondage de tailles décroissantes) que Z' informe sur ζ lorsque Z informe sur ζ et Z' sur Z . Plus précisément : la relation $E Z = \zeta$ est un cas particulier d'une correspondance générale

$$e Z = \zeta$$

qui possède les propriétés suivantes (transitivité).

Soit $Z_v = \zeta, Z_{v-1}, Z_{v-2}, \dots, Z_n, \dots, Z_r$ ($r \geq 1$) les expressions de Z pour toutes les tailles d'échantillons successifs (de $v-1$ à r);

$$(1) \quad e Z_n = \zeta$$

entraîne $e' Z_n = Z_m$ pour tout $m > n$.

(2) Si $e Z_n = \zeta$ et $e' Z_n = Z_n$ on a $e e' Z_n = \zeta$ et l'opérateur $(e e')$ fait partie de la même famille que e .

(3) En particulier, si Z_n est calculable pour $n = r = 1$, on a :

$$Z_1 = f(X, Y, \dots)$$

$$\text{D'où :} \quad e[f(X, Y, \dots)] = \zeta, \quad e'[f(X, Y, \dots)] = Z_m$$

qui donnent les définitions de l'estimateur Z et du paramètre ζ de la population U .

Application.

Reprenons le cas où $\zeta = \bar{x}$, $Z = c \bar{X}$ (Basu; voir IV, I, C ci-dessus).

Soit $Z_{\nu-1} = a \bar{X}$: $\mathcal{E} Z_{\nu-1} = a \zeta$; de proche en proche, il vient $Z_n = a^{\nu-n} \bar{X}$. Pour $n = 1$: $Z_1 = a^{\nu-1} \cdot X$.

De la même façon, on pourrait avoir $Z_{\nu-1} = \bar{X} + b$, d'où

$$Z_n = \bar{X} + (\nu - n)b$$

Rien n'empêche de rendre les coefficients a et b fonctions du rang n de tirage.

Remarques.

1) Soit à estimer ζ par Z , tel que $\mathcal{E} Z = f(\zeta)$. La transitivité n'est possible qu'avec une relation f linéaire.

2) On verra encore au n° 14 que le calcul de la perte d'information n'est possible qu'autant que ζ et chacun des Z ont une expression commune, en ce sens qu'on doit pouvoir calculer une certaine fonctionnelle $\Phi[Z]$ pour chaque Z et en particulier pour ζ .

Par exemple, on ne peut admettre $\zeta = \frac{a+b+c}{3}$, si Z prend l'une des trois valeurs $\sqrt{ab}$, $\sqrt{ac}$, $\sqrt{bc}$ (pour $n = 2$, $\nu = 3$).

12/ - THEOREME 1 bis -

Si $W[Z, \zeta]$ est fonctionnelle de Z et ζ , on a nécessairement $\mathcal{E} W = \mathcal{E} \Phi[Z] - \Phi[\zeta]$, $\Phi[\]$ étant elle-même une fonctionnelle.

En effet $\mathcal{E} W[Z_2, Z_1]$ n'est pas fonctionnelle de ζ ; or ce n'est autre que

$$\mathcal{E} W[Z_2, \zeta] - \mathcal{E} W[Z_1, \zeta]$$

$$\text{d'où : } \mathcal{E} W[Z_2, \zeta] = k_2 - \Phi[\zeta], \quad \mathcal{E} W[Z_2, Z_1] = K_{21} - \Phi[Z_1]$$

$$\text{d'où : } \mathcal{E} W[Z_1, \zeta] = [k_2 - \mathcal{E} K_{21}] + \mathcal{E} \Phi[Z_1] - \Phi[\zeta]$$

qui ne doit pas dépendre du sondage n° 2; comme quoi

$$k_2 = \mathcal{E} K_{21}$$

Remarque.

$W[Z, \zeta]$ peut dépendre non seulement de Z et ζ mais de la distribution de base; le calcul suppose que ces paramètres restent constants quand on remplace $W[Z_2, \zeta]$ par $W[Z_2, Z_1]$. Par exemple, pour l'estimation du ratio, on doit avoir

$$W[Z_2, \zeta] = \left(\frac{A_2 - a}{a} - \frac{B_2 - b}{b} \right)^2; \quad W[Z_1, \zeta] = \left(\frac{A_1 - a}{a} - \frac{B_1 - b}{b} \right)^2$$

$$W[Z_2, Z_1] = \left(\frac{A_2 - A_1}{a} - \frac{B_2 - B_1}{b} \right)^2$$

et non pas
$$W[Z_2, Z_1] = \left(\frac{A_2 - A_1}{A_1} - \frac{B_2 - B_1}{B_1} \right)^2$$

On dira que l'estimateur est uniformément biaisé si on a

$$E\omega[Z_2, Z_1] = 0$$

condition qui entraîne $\mathcal{E}\omega[Z, \zeta] = 0$. Ceci suppose donc que Z_1 est lié à la distribution de Z_2 de façon transitive : c'est le cas pour :

$$E[\alpha(A_2 - A_1) + \beta(B_2 - B_1)] = 0$$

si α et β sont constants.

13/ - Autre remarque.

La nécessité de considérer tantôt une fonction $\Phi(Z)$ et tantôt une fonctionnelle $\phi[Z]$ peut être mise en évidence en considérant la perte d'information attachée à un point.

Considérons le point $M(\bar{X}, \bar{Y})$, estimateur sans biais du point $m(\bar{x}, \bar{y})$ et la perte d'information qui entache cette estimation, soit

$$V(u\bar{X} + v\bar{Y}) = u^2 V\bar{X} + 2uv \text{Cov}(\bar{X}, \bar{Y}) + v^2 V\bar{Y}$$

Considérons les nombres complexes $Z = \bar{X} + i\bar{Y}$ et $\zeta = \bar{x} + i\bar{y}$ affixes de M et m ; avec $\mathcal{E}Z = \zeta$ par définition

$$VZ = V\bar{X} + 2i \text{Cov}\bar{X}\bar{Y} + i^2 V\bar{Y}$$

Passons en coordonnées polaires $Re^{i\theta} = Z$, $re^{i\theta} = \zeta$. La perte

$$VZ = V(Re^{i\theta}) = \mathcal{E}(Re^{i\theta} - re^{i\theta})^2$$

n'est manifestement pas de la forme (où cop signifie perte en commun).

$$\lambda_1 p(R) + 2 \lambda_2 \text{cop}(R, \Theta) + \lambda_3 p(\Theta).$$

On peut isoler dans VZ la perte d'information

$$(\mathcal{E}R^2 - r^2)$$

qui n'est pas une variance (car R n'est pas en général estimateurs sans biais de r).

En tous cas on ne peut y isoler une expression de la forme

$$\mathcal{E}\Phi(\Theta) - \Phi(\theta)$$

Si l'on remarque que $\theta = \text{arc tg } \tau$, $\Theta = \text{arc tg } T$, où τ et T sont les ratios (univers et échantillon), on comprend que l'emploi d'une fonctionnelle soit nécessaire pour (τ, T) aussi bien que pour (θ, Θ) .

Au contraire supposons les unités-échantillons tirées avec des probabilités uniformes pour le rayon vecteur et l'angle polaire; alors on a :

$$\mathcal{E}R = r, \quad \mathcal{E}\Theta = \theta;$$

la perte d'information attachée au point moyen est, disons

$$V(uR + v\Theta)$$

mais les pertes relatives à l'estimation de $\bar{x}$ par $\bar{X}$, de $\bar{y}$ par $\bar{Y}$ ne sont plus séparément de la forme

$$\mathcal{E}\Phi(\bar{X}) - \Phi(\bar{x}), \quad \mathcal{E}\Phi(\bar{Y}) - \Phi(\bar{y}).$$

14/ - Les pertes d'information des estimateurs uniformément biaisés.

THEOREME II bis -

$$a) \text{ Si } \mathcal{E}W = \mathcal{E}\Phi[Z] - \Phi[\zeta]$$

$$\text{on peut poser : } W = \Phi[Z] - \Phi[\zeta] + \Delta[Z, \zeta]$$

$$\text{et on a : } \mathcal{E}\Delta[Z, \zeta] = 0$$

$$b) \text{ Pour que les deux égalités } \mathcal{E}\Delta[Z, \zeta] = 0$$

et

$$E_{\Delta} [Z', Z] = 0$$

entraînent

$$\mathcal{E}_{\Delta} [Z', \zeta] = 0$$

il suffit qu'on ait :

$$\Delta [Z', \zeta] = \lambda \Delta [Z', Z] + \mu \Delta [Z, \zeta]$$

(où λ et μ sont des constantes).

c) Mais, pour que cette relation reste vérifiée lorsque Z vient en ζ , il faut que $\lambda = 1$; et pour que Z' puisse se confondre avec Z il faut que $\mu = 1$.

D'où

$$\Delta [Z', \zeta] = \Delta [Z', Z] + \Delta [Z, \zeta]$$

L'estimateur est uniformément biaisé. Pour cela il suffit qu'on ait :

$$\Delta [Z, \zeta] = \rho \cdot \{ \psi [\zeta] - \psi [Z] \} \quad \text{avec } \rho = \rho [\zeta]$$

THEOREME III bis -

a) Si l'on veut $W [Z, \zeta]$ symétrique, il suffit de poser

$$\begin{aligned} W [Z, \zeta] &= \{ \psi [Z] - \psi [\zeta] \}^2 \\ &= \psi^2 [Z] - \psi^2 [\zeta] + 2 \psi [\zeta] \{ \psi [\zeta] - \psi [Z] \} \end{aligned}$$

d'où

$$\Phi [Z] = \psi^2 [Z]$$

Réciproquement : $W = W^*$ avec $\Delta = \rho (\psi - \psi^*)$

$$W = (\rho - \rho^*) (\psi - \psi^*) / 2$$

impliquent $\Phi^* - \Phi = -(\rho + \rho^*) (\psi - \psi^*) / 2$

$$\equiv \frac{\rho^* \psi^*}{2} - \frac{\rho \psi}{2} + \frac{1}{2} (\rho \psi^* - \rho^* \psi)$$

d'où

$$\rho \psi^* - \rho^* \psi = 0, \quad \frac{\rho^*}{\psi^*} = \frac{\rho}{\psi} = 2 \lambda \quad (\text{constante})$$

d'où

$$\Phi = \lambda \cdot \psi^2$$

b) En dehors du cas symétrique, on peut vouloir exiger seulement de W d'être positive (sauf $W [\zeta, \zeta] = 0$), croissante avec $|Z - \zeta|$. On a

$$W \equiv (\Phi^* - \rho\psi^*) - (\Phi - \rho\psi)$$

Posons $\Omega = \Phi - \psi^2$; faisons $\rho = 2\psi$;

d'où $W = (\psi^* - \psi)^2 + \Omega^* - \Omega$

avec $\Omega = \Omega[\zeta]$, $\Omega^* = \Omega[Z]$

En exigeant $\Omega^* > \Omega$ pour $|Z - \zeta| > 0$, on remplit les conditions voulues.

Conclusion.

Lorsque l'estimateur Z de ζ est biaisé, la perte d'information peut se calculer comme pour un estimateur sans biais, à condition : que Z soit "uniformément" biaisé : $E \omega[Z_2, Z_1] = 0$; qu'on se contente de conditions suffisantes, là où on avait des conditions nécessaires et suffisantes moyennant une régularité suffisante des fonctions.

D'ailleurs en supposant que les fonctionnelles $\Phi[Z]$ ou $\psi[Z]$ sont des fonctions de point (le nombre de dimensions de l'espace Z pouvant même être infini) et en supposant l'existence des dérivées partielles, on peut aller plus loin - on l'a vu aux n°7, 8.

Enfin l'existence des fonctionnelles ω , Φ ou ψ pour tous les Z (y compris ζ) sous-entend en fait que Z et ζ aient une expression commune (sans aller peut-être jusqu'à l'isomorphisme des n°7-8).

III - PERTES AU SENS DU CHAPITRE II -

La perte d'information envisagée ici concerne un plan de sondage bien déterminé et a un certain ordre; il ne lui correspond plus spécialement telle ou telle estimation. On connaît déjà les pertes d'ordre r (et leurs combinaisons linéaires) relatives à un plan de sondage dans une seule urne (Chapitre II). On se propose de montrer que (sous des conditions relativement peu restrictives) il n'existe pas d'autres solutions de l'équation de Chasles que les dites combinaisons linéaires, avec un plan de sondage cette fois très général.

On suppose invariable la structure du sondage, c'est dire que les plans de sondage (π) de la famille (F) sont figurés par les sommets d'un certain treillis.

Quand on sonde une urne unique, l'axe d'information possède un certain point (ω) correspondant à la plus grande perte d'information (non infinie) (c'est-à-dire que ω est le point $n = 1$ pour estimer $\bar{x}$, $n = 2$ pour estimer σ^2 et $\bar{x}$, $n = 3$ pour estimer $\text{cov}(xy)$, σ_x^2 , σ_y^2 , $\bar{x}$ et $\bar{y}$ etc.).

Ce point ω sera l'horizon.

L'horizon d'un treillis de sondage.

Soit une famille F de plans de sondage de treillis

$$(\lambda_1, \lambda_2 \dots \lambda_k)$$

avec $\lambda_i \geq r$. Le point $(r, r, \dots, r)$ en constitue l'horizon.

Exemples.

Sondage stratifié $r \leq n_h \leq v_h$

Sondage à 2 degrés $r \leq m \leq u, \quad r \leq n_i \leq v_i$

Cas d'un treillis découpé en sous-ensembles.

Considérons le cas d'un échantillon stratifié a posteriori - s'il est tiré sans remise et d'ordre r , figuré sur un axe par les points $(n \geq r)$, le point (ϖ) ($n = r$) est l'horizon. En revanche si l'on considère le treillis du sondage stratifié normal dont il dérive (treillis sur lequel on a convenu par exemple que l'estimateur banal, - non stratifié - était affecté aux points tels qu'un n_h au moins est inférieur à r) il existe un sous-ensemble ϖ ($n = r$) et non un point unique ϖ .

Conventions.

Pour simplifier on supposera qu'il existe sur le treillis un horizon et un seul, où aboutissent tous les trajets $[U, \pi, \pi', \dots]$ issus de U (point figurant la connaissance complète); en appelant "trajet" une ligne brisée (empruntant exclusivement les côtés du treillis) le long de laquelle la taille de l'échantillon décroît (c'est-à-dire que l'un des "effectifs" de l'échantillon décroît).

Si d'autres cas se présentent, on ne retiendra que ceux qui peuvent se ramener aux conventions ci-dessus.

Exemple : Sondage à deux phases.

Le treillis $(m, n_1 \dots n_h)$ comprend un axe (représentant la première phase de sondage) sur qui s'appuie une famille de génératrices (m) portant les points figurant la seconde phase. Il n'existe donc pas d'horizon (ϖ). On sort des hypothèses fixées. (Toutefois ce treillis peut être remplacé par un autre, comme on le verra au chapitre VI; et avec des composantes négatives, on pourrait étendre les résultats qui vont suivre).

THEOREME 1 -

Soit I et J deux échantillons du plan de sondage π'' . Soit $g(I, J)$ une fonction symétrique de I et J, nulle quand I et J sont confondus, positive en tout autre cas; g pouvant d'ailleurs dépendre également de la distribution de base (qu'on désigne par U). Si la perte de π'' est de la forme :

$$p(\pi'') = \mathcal{E} g(I, J)$$

et si π' précède π'' sur les trajets $[U\pi']$ (au sens précédent), π' et π'' étant sur le même axe d'information, on a par définition :

$$p(\pi'') = p(\pi') + \mathcal{E} P(\pi''/\pi')$$

d'où

$$p(\pi') = p(\pi'') - \mathcal{E} P(\pi''/\pi')$$

ou

$$p(\pi') = \mathcal{E}_U g(I, J) - \mathcal{E}_U \mathcal{E}_{\pi'} g(I, J)$$

de la forme

$$p(\pi') = \mathcal{E}_U g(I, J) \cdot [1 - \varphi(\pi'/\pi'')]$$

De même, si π précède π' sur les trajets $[U\pi']$, il précède aussi π'' a fortiori; et on a :

$$p(\pi) = \mathcal{E}_U g(I, J) - \mathcal{E}_U \mathcal{E}_{\pi} g(I, J)$$

La condition d'alignement étant (chapitre III, § II) pour les trois points π π' π'' :

$$\mathcal{E} P(\pi''/\pi) - \mathcal{E} P(\pi''/\pi') = \mathcal{E} P(\pi'/\pi),$$

il s'en suit qu'on peut écrire

$$p(\pi) = \mathcal{E}_U g(I, J) [1 - \varphi(\pi/\pi'')]$$

$$p(\pi') - p(\pi) = \mathcal{E}_U g(I, J) [\varphi(\pi/\pi'') - \varphi(\pi'/\pi'')]$$

THEOREME 2 -

Soit I, J, K trois échantillons du plan de sondage π'' et $g(I, J, K)$ une fonction symétrique de I, J, K, nulle quand I, J et K sont tous trois confondus et positive en tout autre cas. Supposons la perte $p(\pi'')$ de la forme

$$p(\pi'') = \mathcal{E}_U g(I, J, K)$$

Si π' précède π'' et π précède π' sur un même axe d'information, on a encore :

$$p(\pi) = \mathcal{E}_U g(I, J, K) \cdot [1 - \omega(\pi/\pi'')]$$

$$p(\pi') - p(\pi) = \mathcal{E}_U g(I, J, K) \cdot [\omega(\pi/\pi'') - \omega(\pi'/\pi'')]$$

Ceci s'étend à un nombre quelconque d'échantillons. Dans ce qui suit, ces extensions à plus de deux échantillons ne seront plus explicitement formulées, encore qu'elles demeurent possibles en principe.

Cas où π'' et σ coïncident.

Supposons en particulier que le plan de sondage π'' soit tel que la perte d'information ne soit plus définie pour aucun trajet $[U\pi]$ où π'' précède π ; ce qui revient à dire que π'' est l'horizon σ . On a donc :

THEOREME 3 -

$$p(\pi) = \mathcal{E}_U g(I, J) [1 - \theta(\pi)] \quad \text{avec } \theta(\sigma) = 0,$$

pour tous les plans de sondage π qui sont sur des axes d'information aboutissant au point σ , pourvu que la perte relative au plan de sondage (σ) soit elle-même de la forme

$$\mathcal{E}_U g(I, J)$$

Combinaisons linéaires de fonctions $g(I, J)$.

THEOREME 4 -

Des fonctions g différentes (représentant chacune une distance entre deux échantillons) peuvent être envisagées simultanément et combinées linéairement, ce qui donne :

$$p(\sigma) = \sum_i \lambda_i \mathcal{E}_U g_i(I, J)$$

$$p(\pi') - p(\pi) = \sum_i \lambda_i \mathcal{E}_U g_i(I, J) [\theta_i(\pi) - \theta_i(\pi')]$$

Les coefficients λ sont des constantes, à moins qu'ils ne dépendent de la distribution de base U . Il n'est pas nécessaire qu'ils soient tous positifs, mais la combinaison $p(\pi)$ doit toujours rester positive (avec $p(U) = 0$).

Remarque. Cas de pertes d'information d'ordres divers.

La difficulté suivante se présentera. On tire des boules d'une urne unique.

D'une part soit $r = 2$. On tire $I(x_h, x_k)$ et $J(x_h, x_k)$ avec quoi on forme par exemple :

$$g_1(I, J) = [(x_h - x_k)^2 - (x_{h'} - x_{k'})^2]^2;$$

d'autre part on choisit $r = 1$, on tire $i(x_i)$ et $j(x_j)$ avec quoi on forme, disons :

$$g_2(i, j) = (x_i - x_j)^4.$$

Il n'est pas question de former $\lambda_1 g_1 + \lambda_2 g_2$.

On adopte comme horizon ϖ le plan de sondage d'effectif $n = 2$; et l'expression g_2 ne saurait être $(x_i - x_j)^4$ mais par exemple :

$$(x_h + x_k - x_{h'} - x_{k'})^4$$

ou bien encore $(x_h - x_{h'})^4 + (x_h - x_{k'})^4 + (x_k - x_{h'})^4 + (x_k - x_{k'})^4$

Telle est la signification du théorème 4.

Réciproque (ou THEOREME 5).

Toute perte $p(\varpi)$ est nécessairement de la forme :

$$p(\varpi) = \sum_u \mathcal{E} g(I, J \dots) [\theta(\mu)^u - \theta(r)]$$

(sous réserve qu'il existe un horizon ϖ).

L'énoncé de ce résultat très important et très général va être précisé au cours de la démonstration.

a) Considérons les arêtes du treillis, ayant leur extrémité en ϖ et leur origine en des points π^0 (qui précèdent immédiatement ϖ sur le trajet $[\nu \dots \varpi]$). Considérons l'un de ces (π^0) et soit E^0 l'un de ses échantillons.

E^0 renferme plusieurs échantillons de (ϖ) , qu'on désignera par $ij \dots$, - et leur donne naissance quand on tire des échantillons de plus en plus petits (en tirant au sort les unités chassées de l'échantillon).

b) La perte d'information supplémentaire, en l'occurrence $P(\varpi/E^0)$, est de la forme $g'(ij \dots)$.

En effet : elle est nulle si tous les $ij \dots$ sont identiques entre eux puis- qu'alors on ne perd rien à en perdre un; elle est positive en tous autres cas; elle dépend symétriquement des $ij \dots$ si ceux-ci sont équiprobables.

c) Par suite l'expression $P(\varpi/\pi^0)$ est de la forme $g'(ij \dots)$

et, en posant :

$$g' \equiv g_1 [\theta(\pi^0) - \theta(\varpi)] \quad \text{avec } \theta(\varpi) = 0, \theta \text{ croissante}$$

on a
$$p(\varpi) - p(\pi^0) \equiv \mathcal{E} g_1 \cdot [\theta(\pi^0) - \theta(\varpi)]$$

Remontons alors l'axe $\pi^0\varpi$ sur toute sa longueur; on aura (théorème 3 ci-dessus) :

$$p(\varpi) - p(\pi) \equiv \mathcal{E} g_1 [\theta(\pi) - \theta(\varpi)]$$

θ étant une fonction croissante quand la taille de l'échantillon augmente. Repérons les plans de sondage π de cet axe par des indices r pour ϖ , $(r+1)$ pour π^0 , $(r+2)$ etc. m (pour π). La suite des m est, soit bornée par un certain μ , soit infinie (tirages avec remise et on arrive alors à un point d'accumulation μ^∞); dans les deux cas on pourra écrire, arrivé à l'extrémité de l'axe opposée à ϖ :

$$p(\varpi) - p(\pi_\mu) \equiv \mathcal{E} g_1 [\theta_\mu - \theta_r]$$

Or on a supposé (par hypothèse) que tout π pouvait être atteint par un trajet issu de U et aussi que tous ces trajets aboutissaient en ϖ . Il existe donc au moins un trajet $[\cup \pi_\mu \varpi]$ qu'on va remonter de π_μ vers U sur un autre axe (le premier ayant été épuisé). Axe qui d'ailleurs comme le premier peut n'être qu'une arête unique.

Le long de cet axe, $p(\pi_\mu) - p(\pi)$ est encore de la forme

$$\mathcal{E} g_2 [\omega_\nu - \omega_s]$$

où ω désigne une fonction croissante quand la taille de l'échantillon augmente : g_2 désigne une fonction symétrique d'un certain nombre d'échantillons de (π_μ) .

De proche en proche, on arrivera ainsi au point U , $p(\varpi) - p(\pi)$ deviendra

$$p(\varpi) - p(U) \equiv p(\varpi)$$

de la forme

$$p(\varpi) \equiv \mathcal{E} g_1 (\theta_\mu - \theta_r) + \mathcal{E} g_2 (\omega_\nu - \omega_s) + \dots + \mathcal{E} g_\lambda (\rho_\tau - \rho_\tau) \quad \text{cqfd}$$

Corollaire.

Toute perte $p(\pi)$ est nécessairement de la forme précédente où θ_μ remplace θ_r . (Démonstration analogue en suivant un trajet $[U \pi]$ à l'envers, en remontant toujours les axes d'information (jusqu'à leur origine par exemple). C'est le théorème annoncé au Ch. II, § IV, Réciproque du Th. 2.

Remarque.

1) La décomposition de $p(\pi)$ ou $p(\omega)$ en λ composantes n'est pas unique. On verra plus loin qu'avec les sondages à λ strates on peut définir strictement λ composantes, tandis que pour les sondages à λ degrés on pourra en définir strictement $2^\lambda - 1$ (si chaque degré de sondage est indépendant des autres).

2) Les fonctions $\theta, \omega, \dots, \rho, \psi$ considérées plus haut ne dépendent pas seulement de la taille des échantillons. Lorsque tous les tirages au sort sont équiprobables, il est bien clair que ces fonctions se déduisent d'un calcul facile d'analyse combinatoire. Mais considérons par exemple une stratification a posteriori en deux strates d'un échantillon bernoullien de taille n : l'expression de $p(n)$ n'est pas aussi simple. Par exemple on aurait (p_1 et p_2 désignant les proportions de la population dans chaque strate).

$$\left\{ \begin{array}{ll} n_1 \neq 0, & n_2 \neq 0 \\ n_1 n_2 = 0 \end{array} \right. \left\{ \begin{array}{l} \bar{X} = p_1 \bar{X}_1 + (1 - p_1) \bar{X}_2; \quad 1 - p_1 = p_2 \\ V \bar{X} = p_1^2 \sigma_1^2 \mathcal{E}\left(\frac{1}{n_1}\right) + p_2^2 \sigma_2^2 \mathcal{E}\left(\frac{1}{n - n_1}\right) \\ \bar{X} = (n_1 \bar{X}_1 + n_2 \bar{X}_2)/n; \quad V \bar{X} = \sigma^2/n. \end{array} \right.$$

On peut poser $\mathcal{E}(1/n_1) = \theta(\infty) - \theta(n)$, mais ne pas oublier qu'une telle expression dépend de p_1 .

Le problème de la forme des fonctions $g(I, J)$.

Soit I, J un couple d'échantillons tiré au sort et $\mathcal{E}$ l'espérance mathématique des tirages avec remise. On sait que

$$V Z = \mathcal{E}(Z - \zeta)^2$$

peut s'écrire, quel que soit le plan de sondage :

$$V Z = \frac{1}{2} \mathcal{E}(Z_i - Z_j)^2 = \mathcal{E} g(I, J)$$

$$\text{avec} \quad g(I, J) = (Z_i - Z_j)^2 / 2 \quad (\text{en supposant } \mathcal{E} Z = \zeta)$$

Ainsi pour une suite de plans de sondage de tailles décroissantes extraite d'une famille F , la perte d'information $V Z$ conserve à toutes les étapes la forme $\mathcal{E} g(I, J)$ avec la même fonction g .

En particulier si la perte est d'ordre 1, l'estimateur $z(X)$ prend pour $n = 1$ les valeurs $z(x_i)$ et $z(x_j)$ sur le couple échantillon x_i, x_j ; et on a :

$$g(i, j) = [z(x_i) - z(x_j)]^2 / 2$$

(en supposant $\mathcal{E}z(X) = \zeta$).

Il n'y a aucune raison pour qu'on ait (avec d'autres fonctions g) :

$$\mathcal{E}_u g(I, J) = \mathcal{E}_u g(i, j) - \mathcal{E}_u \mathcal{E}_\pi g(i, j)$$

et on a peu de chances de se tromper en voyant dans cette conservation de la forme une propriété de la seule variance, parmi toutes les pertes d'information au point de vue du chapitre II.

Mais on n'a pas trouvé le chemin qui conduirait à un résultat aussi général (qui est assez analogue aux problèmes d'agrégation posés par d'autres théories).

IV - PERTE D'INFORMATION D'UN SONDAGE OU D'UNE ESTIMATION ?

1/ - Enoncé du problème.

Dans quel cas est-il possible d'expliquer une perte d'information au sens du Chapitre II (concernant un plan de sondage) en montrant que c'est en fait une perte d'information au sens du Chapitre I (relative à l'estimation d'un paramètre ou d'un groupe de paramètres) ?

Nous sommes loin d'avoir résolu ce problème, mais allons donner quelques résultats fragmentaires le concernant.

2/ - Un principe général.

S'il n'était d'estimations qu'estimations sans biais, il serait possible déjà de poser un principe, d'une application d'ailleurs délicate : Toute perte au sens du Chapitre I peut se mettre sous la forme :

$$\mathcal{E}\Phi(Z) - \Phi(\zeta)$$

et correspond alors à l'estimation sans biais de $\sqrt{\Phi(\zeta)}$ par $\sqrt{\Phi(Z)}$ avec un poids W symétrique, plus des estimations diverses avec poids dissymétrique.

Avec les estimations biaisées, limitées au cas isomorphe comme on l'a fait ici, le même principe fait intervenir des fonctionnelles :

$$\mathcal{E}\Phi[Z] - \Phi[\zeta]$$

Enfin les théorèmes qu'on évoque ici ne sont vérifiés qu'autant que les fonctions ou fonctionnelles auxquelles on a recours possèdent des dérivées pour $\zeta = Z$; l'existence des dérivées de tous ordres est

requis pour le développement en série; une telle condition est peut-être trop restrictive; mais le théorème laisse échapper à coup sûr le cas de $W = |Z - \zeta|$, où la dérivée pour $Z = \zeta$ n'existe pas; or on sait qu'il existe des pertes $\propto |Z - \zeta|$, au moins pour les estimations biaisées du "range" et de la plus grande (ou petite) valeur - Les conditions de régularité nécessaires pour établir les théorèmes généraux laissent donc échapper des cas intéressants.

3/ - Un exemple d'interprétation.

Limitons-nous aux tirages équiprobables dans une urne unique; et étudions le cas de

$$p(v, n) = \mathcal{E}(x_i - x_j)^4 \cdot [\theta(v) - \theta(n)]$$

Cette perte d'information au sens du Chapitre II correspond-elle à une estimation? On précisera ceci ultérieurement.

Calcul de α = $\sum_i \sum_j (x_i - x_j)^4 / v^2$.

Partant de $\frac{1}{v} \sum_{j=1}^v (\xi - x_j)^4 = \xi^4 - 4\xi^3\mu_1 + 6\xi^2\mu_2 - 4\xi\mu_3 + \mu_4$ et en y faisant successivement : $\xi = x_1, x_2 \dots x_v$, sommant et divisant par v , il vient :

$$\alpha = 2(\mu_4 - 4\mu_3\mu_1 + 3\mu_2^2)$$

les μ désignant les moments non-centrés des x_i ; $\mu_1 = \bar{x}$; exprimé avec les moments centrés, il vient :

$$\alpha = 2(\mu_4^0 + 3\sigma^4)$$

Divers essais pour mettre α sous la forme $\mathcal{E}W(z_i, \xi)$ échouent; mais on trouve autre chose.

Première interprétation.

Il était bien naturel qu'on rapprochât α et plus généralement $\sum \sum (x_i - x_j)^{2r} / v^2$, avec le minimum de la fonction

$$y(\xi) = \sum_i (x_i - \xi)^{2r}$$

lorsque ξ décrit l'axe des x ; -minimum qui a lieu pour

$$z(\xi) = \sum_i (x_i - \xi)^{2r-1} = 0$$

équation qui admet une et une seule racine positive, vu que $z'(\xi)$ est nécessairement positive; en particulier, pour $r = 1$, c'est $\bar{x}$.

Dans le cas plus particulier de $r = 2$, désignons par le minimum de y/v , atteint pour $\xi = \xi_0$. On a :

$$y/v = \xi^4 - 4\mu_1\xi^3 + 6\mu_2\xi^2 - 4\mu_3\xi + \mu_4$$

$$0 = [\xi_0^3 - 3\mu_1\xi_0^2 + 3\mu_2\xi_0 - \mu_3]\xi_0$$

$$\eta = \xi_0^4 - 4\mu_1\xi_0^3 + 6\mu_2\xi_0^2 - 4\mu_3\xi_0 + \mu_4$$

d'où
$$\eta = (\mu_4 - \mu_1\mu_3) - 3\xi_0(\mu_3 - \mu_1\mu_2) + 3\xi_0^2(\mu_2 - \mu_1^2)$$

Rapprochons cette expression de celle de

$$\frac{1}{2}\alpha = (\mu_4 - \mu_1\mu_3) - 3\mu_1(\mu_3 - \mu_1\mu_2) + 3\mu_2(\mu_2 - \mu_1^2)$$

Pour qu'elles coïncident, il faut et il suffit qu'on ait :

$$\mu_1^2 = \xi_0^2 = \mu_2$$

ce qui est exclu. Mais la ressemblance formelle entre les deux formules justifie qu'on tienne α pour une distance entre les x_i et un point fictif où ξ_0 est remplacé par μ_1 et ξ_0^2 par μ_2 .

Il est clair que ceci s'étend à toutes les sommes de puissances paires des $(x_i - x_j)$.

Seconde interprétation.

La première interprétation ne permet d'ailleurs pas de retrouver les estimations sans biais de μ_1 et μ_2 ; car la variance simultanée s'écrit

$$\begin{aligned} \mathcal{E}[x_i^2 - \mu_2 + t(x_i - \mu_1)]^2 &= \mathcal{E}(x_i^2 - \mu_2)^2 + 2t\mathcal{E}(x_i^2 - \mu_2)(x_i - \mu_1) + t^2\mathcal{E}(x_i - \mu_1)^2 \\ &= (\mu_4 - \mu_2^2) + 2t(\mu_3 - \mu_1\mu_2) + t^2(\mu_2 - \mu_1^2) \end{aligned}$$

alors qu'on a
$$\frac{\alpha}{2} = (\mu_4 - \mu_2^2) - 4\mu_1(\mu_3 - \mu_1\mu_2) + 4\mu_2(\mu_2 - \mu_1^2)$$

Or il n'est pas possible de faire $t^2 = 4\mu_2$, $t = -2\mu_1$ simultanément, car :

$$\mu_2 \neq \mu_1^2$$

En revanche on peut remplacer t par $(-2x_j)$, où x_j est supposé indépendant de x_i .

Considérons alors la variable (où x_i et x_j sont indépendants)

$$y_k = x_i^2 - 4 x_i x_j + x_j^2$$

on a

$$\mathcal{E} y_k = 2\mu_2 - 4\mu_1^2$$

et

$$\mathcal{V} y_k = \frac{\alpha}{2}$$

1) Il s'agit de tirages bernoulliens.

2) Il s'agit d'une perte d'ordre 2, de sorte que p n'est perte que si la fonction θ est d'ordre 2 (contrairement à ce qu'on avait pu penser, la seule connaissance de α ne permet donc pas de trancher la question).

Remarque.

D'ailleurs le calcul de $\mathcal{V} S^2$ ($nS^2/n - 1$, estimateur de $v\sigma^2/v - 1$) conduit à une composante en α . Les combinaisons linéaires de plusieurs variances conservent les propriétés de récurrence. Bien des informations au sens du Chapitre II ne sont que des composantes des informations au sens du Chapitre I; c'est peut-être la véritable "explication" de leur existence.

4/ - Une méthode de calcul (tirages dans une seule urne).

Considérons une perte d'information d'ordre 1

$$\mathcal{E} g(i, j) \left(\frac{1}{n} - \frac{1}{v} \right)$$

où g est de la forme $g(z_i, z_j)$, avec $z = h(x)$,

$$g > 0, \text{ sauf } g(i, i) = 0.$$

Pour $n = 1$, on imagine un estimateur Z prenant les valeurs z_i ; on ne peut poser

$$W = g(Z, \zeta)$$

En revanche :

THEOREME -

On peut toujours poser

$$W(Z, \zeta) = \mathcal{E}_j g(Z, z_j) - \mathcal{E}_j g(\zeta, z_j) + g(Z, \zeta)$$

Effectivement on vérifie que $\mathcal{E} W(Z, \zeta)$ est identique à

$$\mathcal{E}_{ij} g(z_i, z_j)$$

et que $W(\zeta, \zeta) = 0$, quel que soit ζ .

Choix de ζ .

Lorsque $Z = \zeta + \varepsilon$, $g(Z, \zeta)$ est positif très petit; $g(\zeta + \varepsilon, z_i) - g(\zeta, z_j)$ est du signe de $\varepsilon (\varepsilon - z_j)$; si l'on veut être assuré que W est toujours positif (hors le cas $Z = \zeta$), il faut donc que

$$\mathcal{E}_j g(\zeta, z_j)$$

soit le minimum de $\mathcal{E}_j g(\zeta + \varepsilon, z_j)$. Ceci conduit à choisir ζ tel que $\min \mathcal{E}_j g(Z, z_j) = \mathcal{E}_j g(\zeta, z_j)$.

Remarque.

Tout ceci suppose $n = 1$. Passons au cas où n est quelconque. Par hypothèse

$$p(v, n) + \mathcal{E} P(n, 1) = p(v, 1)$$

$$p(v, n) = p(v, 1) - \mathcal{E} P(n, 1)$$

On vient d'interpréter $p(v, 1) = \mathcal{E} W(Z_1, \zeta)$ et on a de même

$$P(n, 1) = \mathcal{E} W(Z_1, Z_n)$$

Z_n étant défini dans l'urne aléatoire de taille n par la même condition que ζ l'est dans l'univers.

Convention.

On peut alors convenir que $p(v, n)$ représente la distance entre ζ et Z_n , la perte d'information sur ζ quand on connaît Z_n . Ceci est purement conventionnel cette fois, car

$$p(v, n) = \mathcal{E} W(Z_1, \zeta) - \mathcal{E} W(Z_1, Z_n)$$

n'implique pas $p(v, n) = \mathcal{E} W(Z_n, \zeta)$ ni même $\mathcal{E} W[Z_n, \zeta]$.

g est une fonction; $W(Z_1, \zeta)$ et $W(Z_1, Z_n)$ sont des fonctions.

Nous n'avons aucune raison de choisir une fonctionnelle déterminée $W[Z_n, \zeta]$.

Application. $g = |z_i - z_j|^a$.

En écrivant que $\mathcal{E}g(Z, z_j)$ est minimum pour $Z = \zeta$, on retrouve une propriété bien connue de la médiane (ou de l'intervalle médian) lorsque $a = 1$; tandis que, pour $a = 2$ on a (bien entendu) la moyenne. Pour une valeur quelconque de a , il faudrait d'abord discuter l'existence et l'unicité de ζ .

5/ - Etude directe de l'estimation de la médiane.

Entrée en matières.

Pour simplifier on supposera impair l'effectif de la population (soit $2v' + 1$) et celui de chaque échantillon (soit $2n' + 1$).

Partant d'une population entièrement connue, on passera à des échantillons d'effectif décroissant par deux unités à la fois $n = 2v' - 1, 2v' - 3, \dots 5, 3, 1$.

Désignons les médianes par $\zeta = Z_{2v'+1}; Z_{2v'-1}; Z_{2v'-3}; \dots Z_5 Z_3 Z_1$.

En particulier (si les x_i sont tous distincts) :

a) $Z_{2v'-1}$ coïncide avec ζ si les x_i inconnues sont l'un plus grand, l'autre plus petit que ζ ; dans le cas contraire $Z_{2v'-1}$ coïncide avec la valeur x_i immédiatement supérieure, - ou inférieure, - à ζ .

b) Z_1 a une chance sur $(2v' + 1)$ de coïncider avec chacune des valeurs x_i .

c) Les cas intermédiaires sont moins simples; supposons $v = 5$.

Soit $a > b > c > d > e$ ($\zeta = c$) les valeurs x_i considérées. On voit que :

$$\text{pour } n = 3, \quad Z_3 = \begin{cases} c & (\text{probabilité } 4/10) \\ b \text{ ou } d & (\text{probabilité } 3/10) \end{cases}$$

Sauf lorsque la distribution des x_i est symétrique, Z est un estimateur biaisé de ζ .

Perte d'information liée à l'estimation de la médiane.

a) La moyenne et la médiane présentent des analogies importantes. Soit x un point courant de l'axe, $\bar{x}$ la moyenne des x_j et $\tilde{x}$ leur médiane.

$$\begin{aligned} \text{On sait que :} \quad \min \sum |x_j - x| &= \sum |x_j - \tilde{x}| \\ \min \sum (x_j - x)^2 &= \sum (x_j - \bar{x})^2 \end{aligned}$$

On est donc tenté de voir dans $\sum |x_i - \bar{x}|/v$ une perte d'information; en fait cette expression ne possède pas les propriétés nécessaires.

Montrons-le sur l'exemple précédent des cinq boules.

La perte d'information devrait s'évaluer, pour $n' = 1$:

soit directement par : $[|ca| + |cb| + |cd| + |ce|]/5$

soit en deux étapes, par : $[|cb| + |cd|] \frac{3}{10}$ pour $n = 3$

pour $n' = 1$ ($n = 3$), par :

10 combinaisons 3 à 3 de 5 lettres	acd	[ca	+	cd]/30
	ace	[ca	+	ce]/30
	bcd	[cb	+	cd]/30
	...			
	cde	[dc	+	de]/30

L'identification des deux expressions de l'information donne :

$$\frac{2}{10} [|ae| + |bd|] = \frac{3}{10} |bd| + \left[\frac{2}{10} |ae| + \frac{1}{10} |bd| \right]$$

c'est-à-dire $|bd| = 0$

(b, c, d confondus) ce qui est exclu.

b) On est tenté de même de remplacer $(x_i - x_j)^2/2$ par $|x_i - x_j|$ dans l'expression de la perte d'information relative à $\bar{X}$, pour obtenir celle relative à la médiane échantillon $\tilde{X}$, soit :

$$p(v, n) = \frac{\sum |x_i - x_j|}{v(v-1)} \left(\frac{1}{n} - \frac{1}{v} \right) = \frac{1}{n} \frac{\sum \sum |x_i - \bar{x}_j|}{v^2} \cdot \frac{v-n}{v-1}$$

C'est incontestablement une perte d'information. Donnons une expression de

$$\alpha = \frac{1}{v^2} \sum \sum |x_i - x_j|$$

qui rappelle beaucoup celle du paragraphe a, ci-dessus.

Calcul de α .

Désignons par x_n la suite ordonnée :

$$x_1 \leq x_2 \leq \dots \leq x_v$$

Il vient :

$$1 \leq h \leq v-1 \quad \sum_h (x_v - x_h) + (x_h - x_1) = (v-1)(x_v - x_1)$$

$$2 \leq h \leq v-2 \quad \sum_h (x_{v-1} - x_h) + (x_h - x_2) = (v-3)(x_{v-1} - x_2)$$

etc.

etc.

Si v est impair, on termine sur $2 \left[x_{\frac{v+3}{2}} - x_{\frac{v-1}{2}} \right]$

Si v est pair, on termine sur $1. \left[x_{\frac{v+1}{2}} - x_{\frac{v}{2}} \right]$

Pour v impair, on a donc :

$$\alpha = \frac{2}{v^2} \left[(v-1)(x_v - x_1) + (v-3)(x_{v-1} - x_2) + \dots + 2(x_{\frac{v+3}{2}} - x_{\frac{v-1}{2}}) \right]$$

$$\text{Avec} \quad \begin{cases} x_v - x_1 \equiv |x_v - \tilde{x}| + |\tilde{x} - x_1| \\ x_{v-1} - x_2 \equiv |x_{v-1} - \tilde{x}| + |\tilde{x} - x_2| \end{cases} \quad \text{etc.}$$

On peut écrire également :

$$\alpha = \frac{4}{v^2} \sum_h h(x_h - \tilde{x})$$

en désignant cette fois par x_h la suite $\dots x_{-3} x_{-2} x_{-1} \tilde{x} x_1 x_2 x_3 \dots$

Remarque 1.

Affecter aux points x_i des poids ϖ_i symétriques par rapport à leur médiane ne modifie en rien celle-ci; on a donc encore :

$$\min \sum \varpi_i |x_i - x| = \sum \varpi_i |x_i - \tilde{x}|$$

c'est le cas en particulier pour

$$\frac{\varpi_v}{v-1} = \frac{\varpi_{v-1}}{v-3} = \frac{\varpi_{v-2}}{v-5} = \dots = \frac{\varpi_1}{v-1}$$

(le point médian étant affecté d'un poids arbitraire).

Remarque 2.

Si α est la perte d'information $p(v, 1)$ relative à l'estimation de $\tilde{x}$ par l'un quelconque des x_j de la population, $P(n, 1)$ relative à la médiane $\tilde{X}$ d'un échantillon donné lui est isomorphe; c'est-à-dire :

$$A_n = \frac{4}{n^2} \sum h(x_h - \tilde{X}) - \frac{1}{n^2} \sum SS |x_i - x_j|$$

De l'identité

$$\mathcal{E} \left(\frac{n}{n-1} A_n \right) \equiv \frac{v\alpha}{v-1}$$

généralisation du Lemme 2, découle bien :

$$\mathcal{E} P(n, 1) = \alpha \frac{v(n-1)}{n(v-1)} = \alpha \left(1 - \frac{v-n}{n(v-1)} \right)$$

c'est-à-dire : $p(v, n) = \frac{\alpha}{n} \frac{v-n}{v-1}$.

En fait il s'agit d'une convention arbitraire.

Exemple.

$$n = v - 2; \quad p(v, v-2) = \alpha \frac{2}{(v-1)(v-2)}$$

est incompatible avec $p(v, v-2) = \mathcal{E} W(\tilde{X}_{v-2}, \tilde{x})$ où $\tilde{X}_{v-2}$ désigne les valeurs que prend la médiane d'un échantillon d'effectif $n = v - 2$.

$\tilde{X}_{v-2} = x_{\frac{v+1}{2}} = \tilde{x}$	Probabilité = $\left(\frac{v-1}{2} \right) / \frac{v(v-1)}{2} = \frac{v-1}{2v}$
$\tilde{X}_{v-2} = x_{\frac{v-1}{2}} \text{ ou } x_{\frac{v+3}{2}}$	Probabilité = $\left(1 - \frac{v-1}{2v} \right) / 2 = \frac{v+1}{4v}$

Si W était fonction de $\tilde{x}$ et $\tilde{X}$, p serait fonction seulement des trois valeurs centrales de X ; alors que si p est proportionnelle à α , elle dépend de toutes les valeurs de X sauf la médiane.

La contradiction disparaîtrait si l'on convenait que $W[\tilde{X}, \tilde{x}]$ est fonctionnelle de toute la distribution X : nulle quand X prend la valeur $\tilde{x}$, positive autrement, croissante avec $|\tilde{x} - \tilde{X}|$, soit ici :

$$\frac{v+1}{4v} W \left[x_{\frac{v-1}{2}}, \tilde{x} \right] + \frac{v+1}{4v} W \left[x_{\frac{v+3}{2}}, \tilde{x} \right] \equiv \frac{2}{v-1} \frac{\alpha}{v-2}$$

Mais nous ne savons pas partager $(8v\alpha/(v^2-1)(v-2))$ entre les expressions $W \left[x_{\frac{v-1}{2}}, \tilde{x} \right]$ et $W \left[x_{\frac{v+3}{2}}, \tilde{x} \right]$.

6/ - Quelques indications sur le cas où les modules remplacent les carrés.

a) De même qu'on a substitué $|x_i - x_j|$ à $(x_i - x_j)^2$ pour trouver la perte d'information attachée à l'estimation de la médiane de la population par la médiane échantillon, étudions l'expression

$$\delta = |u(x_i - x_j) + v(y_i - y_j)| = |(ux_i + vy_i) - (ux_j + vy_j)|$$

qui remplace

$$V(uX + vY)$$

On peut l'interpréter de plusieurs façons :

Si u et v sont considérés comme des vecteurs de base, δ est le module du vecteur de composantes $(x_i - x_j)$, $(y_i - y_j)$;

Si u et v sont considérés comme des scalaires et (u, v) comme les composantes rectangulaires d'un vecteur,

$$\delta / \sqrt{u^2 + v^2}$$

est la projection du vecteur $(x_i - x_j, y_i - y_j)$ sur la direction (u, v) .

A ce second point de vue

pour $u = 1, v = 0$ on retrouve le problème de la médiane des x_i ;

pour $u = 0, v = 1$ on a celui de la médiane des y_i ;

pour $u^2 + v^2 = 1$, on a celui de la médiane des projections des points (x_i, y_i) sur (u, v) .

Remarque.

Il est exceptionnel que le nuage des points (x, y) ait un point médian M_0 ; ceci suppose le nuage décomposé en deux demi-nuages (M_h) (M'_h) M_h et M'_h étant alignés avec M_0 et de part et d'autre de M_0 (quel que soit h).

Dans le cas général si on fait tourner la direction (u, v) , la projection de tout point (x_i, y_i) a l'occasion d'être une médiane (à tour de rôle).

b) Cas du produit vectoriel.

Faisant pendant au produit scalaire $(\vec{X}, \vec{Y})$ caché dans $V(uX + vY)$, on peut reconnaître le produit vectoriel $(\vec{X} \wedge \vec{Y})$ derrière l'expression

$$\alpha = \mathcal{E}|x_i y_j - x_j y_i|$$

Il est clair que α nul équivaut à $\vec{X}$ et $\vec{Y}$ colinéaire (toutes ses

composantes étant nulles); autrement dit $(x_1, y_1) \dots (x_i, y_i) \dots$ points alignés avec l'origine.

$$\frac{1}{2} |x_i y_j - x_j y_i|$$

est l'aire du triangle de sommets $(0,0)$ (x_i, y_i) (x_j, y_j) ou $OM_i M_j$.

Ainsi α mesure en quelque sorte l'éventail entre les divers rayons vecteurs OM_i autour d'un rayon vecteur central.

$$c) \text{ Sur le cas } |x_i - x_j|^3 \equiv g(i, j).$$

Pour simplifier supposons que la variable x_i prenne quatre valeurs

$$a > b > c > d$$

On aura donc

$$\begin{aligned} \delta &= \frac{2}{4^2} [(a-b)^3 + (a-c)^3 + (a-d)^3 + (b-c)^3 + (b-d)^3 + (c-d)^3] \\ &= \frac{1}{8} [3a^3 + b^3 - c^3 - 3d^3 - 3ab(a-b) - 3ac(a-c) - 3ad(a-d) \\ &\quad - 3bc(b-c) - 3bd(b-d) - 3cd(c-d)] \end{aligned}$$

Le terme $(3a^3 + b^3 - c^3 - 3d^3)$ est analogue au terme $(3a + b - c - 3d)$ qui apparaît quand on traite le cas de la première puissance, c'est

$$\sum_h h(x_h^3 - x_{-h}^3)$$

Les autres termes ont une structure comparable, on reconnaît :

$$\sum x_i x_j |x_i - x_j|$$

Au total δ est la somme des deux termes

$$\frac{2}{\sqrt{2}} \sum |x_i - x_j|^3 \equiv \frac{2}{\sqrt{2}} \sum |x_i^3 - x_j^3| - \frac{6}{\sqrt{2}} \sum |x_i^2 x_j - x_j^2 x_i|$$

Il n'a pas paru qu'en général cette expression pouvait représenter une perte d'information attachée à l'estimation (biaisée) d'un paramètre. Mais il ne paraît pas absurde d'y voir une perte d'information relative à l'estimation simultanée de deux caractéristiques (médiane et rayon vecteur central).

CHAPITRE V

LES SONDAGES A PLUSIEURS DEGRÉS

I - GENERALITES -

L'étude des sondages à deux degrés englobe comme cas limites :

- le sondage en grappe⁽¹⁾ si le 2ème degré disparaît;
- le sondage stratifié si le 1er degré disparaît.

Par sondage à plusieurs degrés on entend ici plus généralement une combinaison quelconque de sondages en degrés, avec stratifications, sous-stratifications, tirages en grappes, probabilités égales ou inégales, avec ou sans remise des boules tirées. Sont exclus les schémas de tirage avec probabilités inégales tels que les unités de sondage ne puissent jamais être prises qu'une seule fois⁽²⁾.

Le découpage est supposé complet et invariable. Enfin les divers tirages au sort sont supposés indépendants entre eux. A chaque tirage, le nombre d'unités de sondage à tirer est prédéterminé, indépendant des tirages précédents, que ce nombre soit fixe ou en proportion fixe de certains paramètres de la population et du sondage. De plus chaque tirage est indépendant des tirages voisins : dans les autres strates et les autres unités de sondage de même rang.

Toutes ces hypothèses sont habituelles en technique de sondage; par exemple établir des compensations entre strates (Goodman et Kish) n'est pas d'usage courant.

(1) Echantillon en grappes - Mêmes formules que pour le sondage à 1 degré, en désignant par σ l'écart-type entre les grappes et non entre les unités.

(2) Car il en résulte des complications que les travaux les plus récents commencent seulement à surmonter (voir T. 5, ch. IV).

II - ECHANTILLON STRATIFIE -

Considérons une population découpée en strates numérotées 1, 2, ...h... Adoptons les notations suivantes :

<u>Strate h</u>	<u>Population</u>
effectif : v_h	$v = \sum_h v_h$
moyenne des x_{hi} : $\bar{x}_h$	$\bar{x} = \sum_h \bar{x}_h / v$
écart-type des x_{hi} : σ_h	estimateur sans biais de $\bar{x}$:
effectif de l'échantillon : n_h	
moyenne des x_{hi} échantillon : $\bar{X}_h$	$\bar{X} = \sum_h \frac{v_h}{v} \bar{X}_h$

La variance est

$$v \bar{X} = \sum_n \left(\frac{v}{v} \right)^2 v(\bar{X}_h)$$

c'est-à-dire

$$= \sum_n \left(\frac{v_h}{v} \right)^2 p(v_h, n_h)$$

Si l'on envisage, toutes choses égales d'ailleurs, un second échantillon d'effectifs n'_h tels qu'on ait (quel que soit h)

$$n'_h \leq n_h$$

on aura :

$$\begin{aligned} v \bar{X}' - v \bar{X} &= \sum_h \left(\frac{v_h}{v} \right)^2 [p(v_h, n'_h) - p(v_h, n_h)] \\ &= \sum_h \left(\frac{v_h}{v} \right)^2 \mathcal{E} P(n_h, n'_h) \end{aligned}$$

Et si l'on pose, symboliquement cette fois :

$$\left. \begin{aligned} v \bar{X} &= p(v, n) \\ v \bar{X}' &= p(v, n') \end{aligned} \right\} \sum_h \left(\frac{v_h}{v} \right)^2 p(n_h, n'_h) = P(n, n')$$

on aura bien la relation :

$$p(v, n') = p(v, n) + \mathcal{E} P(n, n')$$

La variance $v \bar{X}$ est une perte d'information.

Topologie du sondage stratifié.

Figurons tous les estimateurs $\bar{X}$ (ci-dessus) par des points sur un demi-axe, points rangés par ordre de variance croissante; c'est la représentation la plus simple.

Par exemple, on peut porter, à partir de l'origine U , la longueur $\mathcal{V}\bar{X} = (U\bar{X})$; la distance $(\bar{X}\bar{X}')$ de deux points de l'axe est égale à la combinaison des suppléments de perte d'information par strate, tandis que $(\mathcal{V}\bar{X})$ est la combinaison des pertes d'information par strate (combinaisons linéaires à coefficients constants).

Toutefois l'ordre des $\bar{X}$ sur leur axe dépend des grandeurs respectives des paramètres σ_h et $\sigma_{h'}$ de chaque strate. Pour obtenir une figure invariante, on représentera les sondages stratifiés par le grillage (n_h) : à chaque sommet de ce grillage correspond l'estimateur sans biais $\bar{X}$ du plan de sondage correspondant.

Remarque 1.

Si les tirages sont avec remise, les "barreaux" du grillage sont en nombre infini : ils s'accumulent, leurs limites correspondant à la connaissance complète d'une certaine strate.

Remarque 2.

Les "barreaux" du grillage (sur toute leur longueur) sont axes d'information : l'échantillon extrait de toutes les strates, sauf une, restant fixe, la variation de l'effectif de cette seule strate correspond à des tirages dans une urne unique.

Le seul terme de la variance $\mathcal{V}\bar{X}$, qui correspond à cette strate, varie; c'est $\mathcal{V}\bar{X}_h$ (au facteur $(v_h/v)^2$ près); $\mathcal{V}\bar{X}$ vérifie donc bien la condition d'alignement.

Représentation métrique multidimensionnelle. Usage d'une dimension par strate.

Dans l'espace euclidien (E) à L dimensions, (par exemple à 3 dimensions pour 3 strates), on conviendra que le point $\bar{X}$ de coordonnées ξ_h

$$\xi_h = \left(\frac{v_h}{v}\right)^2 \mathcal{V}\bar{X}_h$$

représente l'estimateur $\bar{X}$, l'origine des coordonnées représentant $\bar{X}$

D'où :

$$\mathcal{V}\bar{X} = \sum_h \xi_h$$

Plus généralement, le point $\bar{X}$ représente le plan de sondage (n_h)

Lorsqu'on donne aux divers n_h toutes les valeurs permises 1, 2,

... v_h , l'ensemble des points $\bar{X}$ constitue les "nœuds" ou sommets d'un grillage.

Et lorsqu'on fixe les valeurs de tous les n_h sauf un, le sous-ensemble de points s'aligne sur une demi-droite constituant un "barreau" dudit grillage.

On ne nuit pas à la généralité en se limitant ici à 3 strates.

a) Soit $\bar{X}$ et $\bar{X}'$ deux estimateurs représentés par deux points du même "barreau", - disons que :

$$n_1 > n'_1; \quad n_2 = n'_2; \quad n_3 = n'_3.$$

On aura :

$$v \bar{X}' - v \bar{X} = \mathcal{E}(\bar{X}' - \bar{X})^2 = \left(\frac{v_1}{v}\right)^2 \frac{v_1 \sigma_1^2}{v_1 - 1} \left(\frac{1}{n'_1} - \frac{1}{n_1}\right)$$

b) Supposons à présent $n_1 > n'_1$; $n_2 > n'_2$; $n_3 = n'_3$. On peut définir un point intermédiaire $\bar{X}''$ (n_1, n'_2, n_3) de façon que :

$$v \bar{X}' - v \bar{X}'' = \mathcal{E}(\bar{X}' - \bar{X}'')^2 = \left(\frac{v_1}{v}\right)^2 \frac{v_1 \sigma_1^2}{v_1 - 1} \left(\frac{1}{n'_1} - \frac{1}{n_1}\right)$$

$$v \bar{X}'' - v \bar{X} = \mathcal{E}(\bar{X}'' - \bar{X})^2 = \left(\frac{v_2}{v}\right)^2 \frac{v_2 \sigma_2^2}{v_2 - 1} \left(\frac{1}{n'_2} - \frac{1}{n_2}\right)$$

$$v \bar{X}' - v \bar{X} = \mathcal{E}(\bar{X}' - \bar{X}'')^2 + \mathcal{E}(\bar{X}'' - \bar{X})^2 = \mathcal{E}(\bar{X}' - \bar{X})^2$$

Ici les segments $\bar{X}\bar{X}''$ et $\bar{X}''\bar{X}'$ sont parallèles aux axes; dans le cas général, c'est tout une ligne polygonale $\bar{X}\bar{X}''$, $\bar{X}'\bar{X}''$, ..., $\bar{X}''\bar{X}'$, dont chaque segment est parallèle à un axe de coordonnées, qu'il y aurait lieu de considérer.

Une confusion à éviter.

On évitera de confondre l'espace (E) considéré ci-dessus avec l'espace (A) dont on fait aussi usage à propos de sondage. Tout aléatoire y est figuré par un vecteur. Deux aléatoires indépendants sont figurés par deux vecteurs orthogonaux de (A), l'espérance mathématique du carré de la différence entre deux vecteurs correspond au carré de la distance dans (A).

Par exemple, $\bar{X}$ et $\bar{X}'$ étant définis comme ci-dessus, on a :

$$\mathcal{E}(\bar{X}' - \bar{X})^2 = \mathcal{E}(\bar{X}' - \bar{X}'')^2 + \mathcal{E}(\bar{X}'' - \bar{X})^2$$

$$\mathcal{E}(\bar{X}' - \bar{X}) \cdot (\bar{X} - \bar{x}) = 0$$

relation qui correspond à l'orthogonalité des vecteurs $\bar{x}\bar{X}$ et $\bar{X}\bar{X}'$ dans l'espace (A) évoqué, et non dans l'espace euclidien (E).

Bien au contraire, on a vu dans l'exemple (a) que $\bar{X}\bar{X}'$ pouvait être parallèle à un axe, alors que $\bar{x}\bar{X}$ n'était pas forcément parallèle à l'autre axe, et d'ailleurs sans qu'on ait exigé que les axes de coordonnées de (E) soient rectangulaires.

On peut en particulier en déduire le paradoxe suivant. Posons :

$$X^* = \lambda \bar{X} + (1 - \lambda) \bar{X}'$$

estimateur sans biais de $\bar{x}$ au même titre que $\bar{X}$ et $\bar{X}'$; dans l'espace (A) il ne fait aucun doute que le point représentant cette combinaison linéaire de $\bar{X}$ et $\bar{X}'$ est sur le segment joignant les points $\bar{X}$ et $\bar{X}'$. On pourrait donc croire que, dans l'espace euclidien (E) il en serait de même. Ce serait inexact.

Le calcul montre immédiatement que :

$$\mathcal{E}(X^* - \bar{X})^2 = (1 - \lambda)^2 \mathcal{E}(\bar{X}' - \bar{X})^2$$

$$\mathcal{E}(X^* - \bar{X}')^2 = \lambda^2 \mathcal{E}(\bar{X} - \bar{X}')^2$$

et il est donc impossible qu'on ait :

$$\text{distance } \bar{X}\bar{X}' = \text{distance } \bar{X}X^* + \text{distance } X^*\bar{X}'$$

lorsque $\bar{X}$ et $\bar{X}'$ sont (comme en a) sur un barreau du grillage.

Sondages stratifiés particuliers.

a) Sondages représentatifs.

De l'ensemble des points $\bar{X}$ ou (n_h) on va extraire les points $(f \vee_h)$ sous-ensemble à un seul paramètre f ; dans l'espace (E), les ξ_h sont fonctions linéaires de $1/f$ et les points sont alignés.

$$\xi_h = a_h \left(\frac{1}{f} - 1 \right)$$

Lorsque $f \vee_h$ n'est pas entier, on peut convenir de rejeter ou non les points. On retiendra que le sondage "représentatif" est représenté par une division sur un axe, homothétique de celle du sondage simple.

b) Sondage "optimum" au sens de Neyman et Yates.

On sait qu'en adoptant des n_h proportionnels aux $v_h \sigma_h / \sqrt{C_h}$, on rend la variance $\mathcal{V}\bar{X}$ minimum à coût constant $\sum C_h n_h = C$.

Les points (n_h) correspondants (après arrondissement des n_h) se trouvent sensiblement eux aussi sur un axe d'information :

$$\xi_h = b_h - \frac{1}{C} b'_h$$

On sait effectivement que $\mathcal{V}\bar{X}$ a pour expression :

$$\text{sondage représentatif : } \sum \left(\frac{v_h}{v} \right)^2 \frac{\sigma_h^2}{v_h - 1} \left(\frac{1}{f} - 1 \right);$$

$$\text{sondage "optimum" : } \sum \left(\frac{v_h}{v} \right) \frac{\sigma_h^2}{v_h - 1} - \frac{1}{C} \left[\sum \frac{v_h}{v} \sigma_h \sqrt{\frac{v_h C_h}{v_h - 1}} \right]$$

III - SONDAGE A 2 DEGRES A 2 PARAMETRES, AVEC TIRAGES EQUIPROBABLES -

On a indiqué déjà (Ch. III, § II) que le sondage à 2 degrés et 2 paramètres $(m, \bar{n})$ (avec $v_i = \bar{v}$) correspondait au treillis le plus simple : le grillage $(m, \bar{n})$ dont les sommets sont repérés par les entiers $(m, \bar{n})$, $1 \leq m \leq \mu$, $1 \leq \bar{n} \leq \bar{v}$.

Le point $(\mu, \bar{v})$ représente la connaissance parfaite; tandis que le point $(m, \bar{n})$ représente les sondages de taille $(m, \bar{n})$.

$$\text{Posons : } \bar{X} = \sum_i S_i x_{ij} / m \bar{n}; \quad \bar{\bar{x}} = \sum_i \sum_j x_{ij} / \mu \bar{v}$$

$$\bar{x} = \sum_i \bar{x}_i / m, \quad \text{avec} \quad \bar{\bar{x}}_i = \sum_j x_{ij} / \bar{v}$$

$$\bar{\bar{X}} = \sum_i \bar{\bar{X}}_i / \mu, \quad \text{avec} \quad \bar{\bar{X}}_i = \sum_j x_{ij} / \bar{n}$$

Attachons $\bar{X}$ à $(m, \bar{n})$ et $\bar{\bar{x}}$ à $(\mu, \bar{v})$, $\bar{x}$ à $(m, \bar{v})$ et $\bar{\bar{X}}$ à $(\mu, \bar{n})$.

Nous avons montré ailleurs (T. 4) que, pour tous les sondages à deux degrés (indépendants l'un de l'autre), on avait :

$$\begin{aligned} \mathcal{V}\bar{X} &= \mathcal{E}(\bar{X} - \bar{\bar{x}})^2 = \mathcal{E}(\bar{X} - \bar{x})^2 + \mathcal{E}(\bar{x} - \bar{\bar{x}})^2 \\ &= \mathcal{E}(\bar{X} - \bar{\bar{X}})^2 + \mathcal{E}(\bar{\bar{X}} - \bar{x})^2 \end{aligned}$$

c'est-à-dire :

$$\mathcal{E}(\bar{X} - \bar{x}) \cdot (\bar{x} - \bar{\bar{x}}) = 0$$

$$\mathcal{E}(\bar{X} - \bar{\bar{X}}) \cdot (\bar{\bar{X}} - \bar{x}) = 0$$

A la première décomposition de $\mathcal{V}\bar{X}$ correspond une première famille d'axes :

$$\mathcal{E}(\bar{x} - \bar{\bar{x}})^2 = \frac{\sigma_o^2}{m} \frac{\mu - m}{\mu - 1} = \text{constante}$$

donc

$$m = \text{constante; } \bar{n} \text{ variable (verticales)}$$

avec

$$\mathcal{V}\bar{X} - \mathcal{E}(\bar{x} - \bar{\bar{x}})^2 = \frac{1}{m\mu} \sum_i \frac{\bar{v} \sigma_i^2}{\bar{v} - 1} \left(\frac{1}{\bar{n}} - \frac{1}{\bar{v}} \right)$$

A la deuxième décomposition de $\mathcal{V}\bar{X}$ correspond une deuxième famille d'axes

$$\left\{ \begin{array}{l} \mathcal{E}(\bar{X} - \bar{\bar{x}})^2 = \frac{1}{\mu^2} \frac{\sum \sigma_i^2}{\bar{n}} \frac{\bar{v} - \bar{n}}{\bar{v} - 1} = \text{constante} \\ \bar{n} = \text{constante; } m \text{ variable (horizontales)} \end{array} \right.$$

$$\text{avec } \mathcal{V}\bar{X} - \mathcal{E}(\bar{\bar{X}} - \bar{\bar{x}})^2 = \frac{\mu \sigma_o^2}{\mu - 1} + \frac{1}{\mu} \sum \frac{\bar{v} \sigma_i^2}{\bar{v} - 1} \left(\frac{1}{\bar{n}} - \frac{1}{\bar{v}} \right) \left(\frac{1}{m} - \frac{1}{\mu} \right)$$

La formule de récurrence est valable pour $\mathcal{V}\bar{X}$ à condition de compter le trajet $\bar{X}\bar{X}'$ le long d'axes d'information et jamais directement entre deux points qu'un axe ne relie pas. Entre les points $(m, \bar{n})$ et $(m', \bar{n})$, $m > m'$, $\bar{n} > \bar{n}'$ on doit suivre un trajet en ligne brisée empruntant les côtés du grillage.

Remarque.

On a sur les premiers axes

$$\vartheta(v) - \vartheta(\pi) = \left[\frac{\bar{v} - 1}{\bar{v}} - \frac{\bar{n} - 1}{\bar{n}} \right] = \frac{1}{\bar{n}} - \frac{1}{\bar{v}}$$

et sur les seconds

$$\vartheta(v) - \vartheta(\pi) = \left[\frac{\mu - 1}{\mu} - \frac{m - 1}{m} \right] = \frac{1}{m} - \frac{1}{\mu}$$

Représentation métrique

Dans l'espace euclidien, considérons le point de coordonnées (α, β, γ)

$$\alpha = \mathcal{E}(\bar{x} - \bar{\bar{x}})^2; \quad \beta = \mathcal{E}(\bar{\bar{X}} - \bar{\bar{x}})^2; \quad \gamma = \left(\frac{\mu}{m} - 1\right)\beta$$

On a par ailleurs :

$$\alpha = \frac{\sigma_o^2}{\mu - 1} \left(\frac{\mu}{m} - 1\right)$$

donc :

$$\frac{\sigma_o^2}{\mu - 1} \gamma = \alpha \beta$$

équation d'un parabolofde hyperbolique P.H. dont les deux familles de génératrices correspondent respectivement à $\alpha = \text{constante}$ et $\beta = \text{constante}$; c'est-à-dire se projettent suivant le grillage $(m, \bar{n})$ sur $(\alpha \beta)$.

Double décomposition de $\mathcal{V}\bar{X}$.

$$\mathcal{V}\bar{X} = \alpha + (\beta + \gamma) = \mathcal{E}(\bar{x} - \bar{\bar{x}})^2 + \frac{\mu}{m} \left[\frac{1}{\mu^2} \frac{\Sigma \sigma_i^2}{\bar{n}} \frac{\bar{v} - \bar{n}}{\bar{v} - 1} \right]$$

$$= \mathcal{E}(\bar{x} - \bar{\bar{x}})^2 + \mathcal{E}(\bar{\bar{X}} - \bar{\bar{x}})^2$$

$$\mathcal{V}\bar{X} = (\alpha + \gamma) + \beta = \mathcal{E}(\bar{\bar{X}} - \bar{\bar{x}})^2 + \mathcal{E}(\bar{\bar{X}} - \bar{\bar{x}})^2$$

car $\mathcal{E}(\bar{X} - \bar{\bar{X}})^2 = \mathcal{E}(\bar{x} - \bar{\bar{x}})^2 + \mathcal{E}[(\bar{X} - \bar{x})(\bar{\bar{X}} - \bar{\bar{x}})]^2 = \alpha + \gamma$

En résumé les plans de sondage à deux degrés, dans le cas où on a :

$$n_i = \bar{n}, \quad v_i = \bar{v},$$

sont représentés par les nœuds d'un treillis dessiné sur un certain parabolofde hyperbolique (P.H.), treillis dont la projection sur le plan des (α, β) est un grillage.

Les côtés des mailles du treillis sont tous portés par des génératrices (de l'un ou l'autre système) de P.H.

La perte d'information est la somme $(\alpha + \beta + \gamma)$ des trois coordonnées des points figuratifs.

Les points situés sur des plans parallèles, d'équation :

$$\alpha + \beta + \gamma = \text{constante}$$

représentent des sondages donnant d'égales pertes d'information.

Autre métrique.

Représentons le même phénomène avec des coordonnées α' , β' , γ' égales à $\sqrt{\alpha}\sqrt{\beta}\sqrt{\gamma}$. C'est même l'idée la plus naturelle, quand on a l'habitude de considérer $\mathcal{E}(\bar{X} - \bar{x})^2$, etc. comme les carrés de certaines distances, et les relations

$$\begin{aligned}\mathcal{E}(\bar{X} - \bar{x})^2 &= \mathcal{E}(\bar{X} - \bar{\bar{X}})^2 + \mathcal{E}(\bar{\bar{X}} - \bar{x})^2 \\ &= \mathcal{E}(\bar{X} - \bar{x})^2 + \mathcal{E}(\bar{x} - \bar{x})^2\end{aligned}$$

comme deux expressions du théorème de Pythagore.

Alors la perte d'information est : $\alpha'^2 + \beta'^2 + \gamma'^2$ c'est-à-dire le carré de la distance de $\bar{x}$ à $\bar{\bar{X}}$.

Il est à noter que (P.H.) se transforme en un autre parabolofde hyperbolique P.H' d'équation :

$$\frac{\sigma_0}{\sqrt{\mu - 1}} \gamma' = \alpha' \beta'$$

toutes les propriétés relatives aux treillis, grillage et génératrices rectilignes se conservent. Mais les lieux des points d'égales pertes d'information sont les sphères de centre $\bar{\bar{X}}$.

Remarque.

On évitera de confondre P.H. ou P.H' et l'espace (A) des sondages où, $\bar{X}_1$ et $\bar{X}_2$ étant représentés par deux points d'une même génératrice de P.H. ou P.H', on aurait :

$$(\bar{\bar{x}} \bar{X}_2)^2 = (\bar{\bar{x}} \bar{X}_1)^2 + (\bar{X}_1 \bar{X}_2)^2$$

Il serait absolument vain de vouloir représenter cet espace dans l'espace à trois dimensions ; chaque génératrice du P.H. représente elle-même un espace euclidien à v dimensions.

IV - LES SONDAGES A PLUSIEURS DEGRES EN GENERAL -

1/ - Tirages équiprobables : Cas de deux degrés de sondage.

La double décomposition de la variance a été démontrée dans le cas général, à la seule condition d'écrire :

$$S(\bar{X} - \bar{x})^2 = \frac{1}{m\mu} \sum_i \frac{\sigma_i^2}{n_i} \frac{v_i - n_i}{v_i - 1}$$

et en supposant exhaustifs les tirages élémentaires.

Au lieu de deux paramètres ($m, \bar{n}$), il y a $(\mu + 1)$ paramètres (m, n_i), car on suppose les n_i définis pour toutes les unités du 2ème degré, tirées ou non.

Si m ne change pas, la variation des n_i a le même effet que sur un sondage stratifié, de sorte que $\mathcal{V}\bar{X}$ se comporte comme une perte d'information au sens du chapitre I, représentable sur un axe. A chaque m correspond un tel axe, et leur ensemble peut être figuré par des barreaux parallèles empruntés à un grillage. Pourtant chaque axe symbolise un grillage à μ dimensions comme pour un sondage stratifié.

Si l'on fait varier m tout seul sans toucher aux n_i , la seconde décomposition de la variance entre en jeu; et il existe une seconde famille d'axes (n_i) représentée par les seconds barreaux du grillage, croisés avec les premiers. Ces axes ne sont d'ailleurs pas ordonnés, de même que l'ensemble des points (n_i).

Pour passer d'un plan de sondage à un autre (d'effectif plus réduit) on se déplacera sur une génératrice d'un système, puis sur une génératrice de l'autre système; et on aura encore le droit de dire que la variance est une perte d'information.

Représentation euclidienne métrique.

α n'est pas modifié et β ne l'est guère :

$$\beta = \frac{1}{\mu} \sum_i \frac{\sigma_i^2}{n_i} \frac{v_i - n_i}{v_i - 1}$$

On a encore :

$$\gamma = \frac{\mu - m}{m} \beta$$

Conclusion.

On retrouve le même P.H. qu'en II, et la même famille (dénombrable) de génératrices $m = \text{constante}$; en revanche sur le deuxième faisceau de génératrices, la substitution des n_i à n augmente très notablement le nombre de génératrices du P.H. réellement utilisées comme axes d'information.

2/ - Extension à un nombre quelconque de degrés de sondage.

On sait que la variance d'un sondage à trois degrés et plus se décompose (et de plusieurs façons) comme celle du sondage à deux degrés. Il est possible de représenter les sondages à d degrés par les nœuds d'un grillage généralisé à d dimensions.

Sur un barreau quelconque du grillage, $\forall \bar{X}$ satisfait à la condition de récurrence. Pour comparer deux sondages quelconques, il faudra se déplacer de proche en proche le long des barreaux. Si les effectifs de l'un sont tous au moins égaux à ceux de l'autre, on peut alors écrire la condition de récurrence.

3/ - Cas où les tirages élémentaires sont bernoulliens.

On sait que la double décomposition de la variance du sondage à deux degrés est encore valable avec des tirages bernoulliens. La représentation par P.H. est encore valable :

a) en remplaçant $(\mu - 1)$ par μ dans l'équation du P.H. ;

b) en faisant intervenir une infinité dénombrable de génératrices du P.H. de chaque famille ($1 \leq \bar{n}$, $1 \leq m$) et non plus un nombre fini de génératrices.

Les grillages comprennent une infinité dénombrable de barreaux des deux systèmes.

On peut accroître le nombre de degrés de sondage, - ou employer les sondages bernoulliens pour certains degrés et les sondages exhaustifs pour d'autres degrés.

4/ - Cas où il existe des strates et des sous-strates.

A condition d'adapter les axes utilisés, rien n'empêche de compliquer le plan de sondage avec des strates (au 1er degré), des sous-strates (au 2ème degré), etc.

5/ - Cas où les tirages ne sont plus équiprobables.

Limitons-nous au cas le plus courant où les probabilités de tirage au 1er degré sont proportionnelles à la taille des unités primaires de sondage et où une seule unité primaire est tirée; au second degré, les tirages sont supposés équiprobables.

On sait que la double décomposition de la variance est valable avec des probabilités de tirage inégales, à condition de modifier légèrement les définitions de $\bar{X}$ et $\bar{x}$ (moyennes pondérées pour $\bar{X}$ et $\bar{x}$, simples pour $\bar{X}$ et $\bar{x}$).

On peut donc étendre la théorie du grillage et aussi celle du parabolofde hyperbolique.

$\bar{X}$ suit la formule de récurrence le long des axes d'information, et cette formule s'étend à un trajet en ligne brisée empruntant les axes ou génératrices; ceci suppose que la taille de l'échantillon s'est réduite (c'est-à-dire qu'aucun des effectifs n'augmente quand certains diminuent).

V - APPLICATION A L'ETUDE DU SONDAJE A 3 DEGRES (ET A 3 PARAMETRES) -

1/ - Décomposition de la variance en 7 composantes.

On abandonnera les notations précédentes (Réf.T.4) pour alléger. Si on désigne par X_{III} ou III l'estimateur sans biais de X_{000} ou 000, la variance de cet estimateur s'écrit :

$$\begin{aligned} V &= E(X_{III} - X_{000})^2 = E(III - 000)^2 \\ &= E(III - 0II)^2 + E(0II - 00I)^2 + E(00I - 000)^2 \\ &\text{etc.} \end{aligned}$$

Elle peut donc prendre six expressions différentes, qu'il est commode de représenter comme suit (avec 12 composantes) :

$$\begin{array}{ll} a + B + \gamma & a + C' + \beta \\ b + C + \alpha & b + A' + \gamma \\ c + A + \beta & c + B' + \alpha \end{array}$$

Toutefois on a :

$$a + B = b + A' = a + b + (ab) = V - \gamma$$

$$b + C = c + B' = b + c + (bc) = V - \alpha$$

$$c + A = a + C' = c + a + (ca) = V - \beta$$

Ainsi la variance V s'exprime à l'aide de 9 composantes seulement : $a, b, c, (ab), (bc), (ca), \alpha, \beta, \gamma$.

De même, il est clair que :

$$V - a = B + \gamma = C' + \beta = B + C' + (BC')$$

$$V - b = C + \alpha = A' + \gamma = C + A' + (CA')$$

$$V - c = A + \beta = B' + \alpha = A + B' + (AB')$$

En combinant ces résultats, il vient :

$$\begin{aligned} V &= a + [b + (ab)] + [c + (ca)] + (BC') \\ &= a + b + c + (ab) + (ca) + (bc) + (BC') - (bc) \end{aligned}$$

et deux formules analogues; d'où il suit que :

$$(BC') - (bc) = (CA') - (ca) = (AB') - (ab) = z$$

et trois formules semblables, d'où trois expressions symétriques :

$$(B'C) - (\beta\alpha) = (C'A) - (\gamma\alpha) = (A'B) - (\alpha\beta)$$

On posera :

$$z = (abc)$$

en remarquant qu'avec le sondage à 2 degrés, les 3 composants de la variance V sont représentés par le symbole :

$$a + b + (ab) = (1 + a) \cdot (1 + b) - 1$$

et qu'avec le sondage à 3 degrés, le symbole à employer est, cette fois :

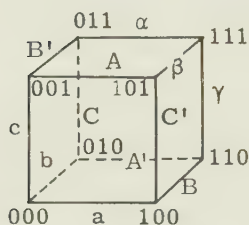
$$(1 + a) \cdot (1 + b) \cdot (1 + c) - 1 = a + b + c + (ab) + (bc) + (ca) + (abc)$$

2/ - Théorème.

Les 7 composantes précédentes de V sont des variances (et par conséquent sont positives).

En effet :

(1) (ab) (bc) (ca) sont par définition des covariances :



$$(ab) = B - b = A' - a$$

$$= E (110 - 100)^2 - E(010 - 000)^2$$

$$= E (110 - 010)^2 - E(100 - 000)^2$$

$$= E [(110+000-100-010)(110+010-100-000)]$$

$$= E [(110+000-100-010)(110+100-010-000)]$$

(2) Mais la théorie du sondage à 2 degrés⁽¹⁾ établit que (ab) est aussi une variance

$$(ab) = E(110 - 100 - 010 + 000)^2$$

Ceci résulte du choix d'estimateurs sans biais 110, 100 et 010,

qui entraîne l'orthogonalité de (110 - 100) et (100 - 000)

et celle de (110 - 010) et (010 - 000)
(comme sur le cube de la figure ci-dessus)

$$E [(110 - 100). (100 - 000)] = E [(110 - 010). (010 - 000)] = 0$$

avec

$$E [(100 - 000). (010 - 000)] = E [(110 - 100). (110 - 010)] = 0$$

D'où

$$E [(110-100). (010-000)] = E [(110-010+010-000+000-100)(010-000)]$$

$$= 0 + E (010 - 000)^2 + 0$$

D'où :

$$E(110-100-010+000)^2 = E(110-100)^2 + E(010-000)^2 - 2E [(110-100)(010-000)]$$

$$= E(110-100)^2 + E(010-000)^2 - 2E(010-000)^2$$

$$= E(110-100)^2 - E(010-000)^2$$

c. q. f. d.

(1) Cf. Annexe de l'Etude théorique n°6 de l'I. N. S. E. E. (Réf. T. 4).

(3) Or il est clair que (abc) joue vis-à-vis de (ab) (bc) et (ac) le même rôle que (ab) vis-à-vis de a et b . Cette différence de deux variances est encore une variance du fait de l'orthogonalité (c'est-à-dire de l'absence de biais des estimateurs).

C. Q. F. D.

Remarque.

Par exemple, dans les cas d'ordre $r = 1$, avec des tirages sans remise équiprobables, on sait que (ab) est de la forme :

$$(ab) = B - b = \frac{H}{\ell} - \frac{H}{\lambda} = H \left(\frac{1}{\ell} - \frac{1}{\lambda} \right) = M$$

De même :

$$(abc) = (AB') - (ab) = \frac{M}{n} - \frac{M}{v} = M \left(\frac{1}{n} - \frac{1}{v} \right)$$

Si l'on se souvient que :

$$b = \frac{H}{\lambda} = \frac{K}{\lambda} \left(\frac{1}{m} - \frac{1}{\mu} \right)$$

On arrive finalement à :

$$(abc) = K \left(\frac{1}{\ell} - \frac{1}{\lambda} \right) \cdot \left(\frac{1}{m} - \frac{1}{\mu} \right) \cdot \left(\frac{1}{n} - \frac{1}{v} \right)$$

3/ - Interprétation géométrique.

a) Dans l'espace à 3 dimensions, un parallélépipède de sommets (000) , (001) , ... (111) , dont les 12 arêtes s'appellent $a b c$, $A B C$, $A' B' C'$, $\alpha \beta \gamma$, donne du plan de sondage une représentation topologique valable. La famille des plans de sondage est représentée par un grillage à 3 dimensions.

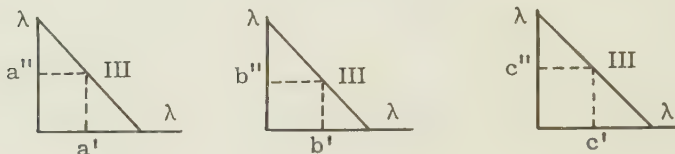
b) La représentation métrique nécessite au contraire un espace euclidien dont le nombre de dimensions n'est pas évident. On pourrait en prendre 9.

3 pour $a b c$;

3 pour (bc) (ca) (ab) ou $a' b' c'$;

3 pour $\alpha - a$, $\beta - b$, $\gamma - c$ ou $a'' b'' c''$

c) Mais considérons les trois plans de coordonnées $a' a''$, $b' b''$, $c' c''$.



Posons $V - (a + b + c) = \lambda$. On a $\lambda = a' + a'' = b' + b'' = c' + c''$ (voir la figure) c'est-à-dire que le point x_{III} est astreint à rester sur une multiplicité linéaire à 4 paramètres dans le sous-espace $(a' b' c' a'' b'' c'')$ à 6 dimensions.

Il est donc naturel de chercher une représentation à $9 - 2 = 7$ dimensions où V soit la somme des 7 coordonnées. C'est le cas pour

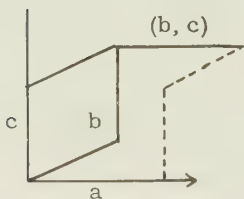
$$V = a + b + c + (ab) + (bc) + (ac) + (abc);$$

cette décomposition (parmi bien d'autres) n'est celle retenue que parce que le terme résiduel (abc) est une variance lui aussi.

4/ - Les trois familles d'axes d'information.

Il est clair qu'il y a symétrie entre tous les degrés de sondage, c'est-à-dire ici entre $(a' a'')$ $(b' b'')$ et $(c' c'')$.

On fera varier simplement le paramètre du 1er degré pour obtenir la 1ère famille de génératrices rectilignes; il lui correspondra une famille analogue pour chacun des autres degrés de sondage.



Faisons varier a seulement, b, c et (bc) restent constantes, pendant que a (ab) (ac) (abc) varient. Les projections des points x_{III} sur le sous-espace $a, b, c, (bc)$, sont alignées, mais on ignore ce qui se passe dans le sous-espace $(ab), (ac), (abc)$. Explicitons les composantes. Par exemple avec des tirages sans remise équiprobables, on a :

$$(ab) = k \left(\frac{I}{n} - \frac{I}{v} \right),$$

$$(ac) = k \left(\frac{I}{n} - \frac{I}{v} \right),$$

$$\text{avec } a = h \left(\frac{I}{n} - \frac{I}{v} \right)$$

$$(abc) = k'' \left(\frac{I}{n} - \frac{I}{v} \right),$$

soit 4 fonctions linéaires de $(1/n)$; d'où l'axe (b, c) ; d'où :

THEOREME -

Si la perte d'information est la variance, il existe trois familles d'axes d'information pour le sondage à 3 degrés; leur représentation métrique euclidienne demande un espace à 7 dimensions.

Corollaire.

Pour le sondage à d degrés, il y a d familles d'axes, dans l'espace euclidien à $2^d - 1$ dimensions.

Remarque. Changement de métrique.

Lorsqu'on emploie une perte d'information plus générale que la variance, la décomposition de cette perte en ses $(2^d - 1)$ composantes est invariante; et même tous les facteurs tels que $\frac{1}{\ell} - \frac{1}{\lambda}$, $\frac{1}{m} - \frac{1}{\mu}$, $\frac{1}{n} - \frac{1}{v}$, $\Phi(v) - \Phi(n)$ ne sont pas modifiés.

V - SUR LES PROBLEMES DE REPARTITION OPTIMUM -

1/ - Généralités.

La répartition optimum (au sens de Neyman) de l'échantillon est celle qui (pour un effectif ou un coût de sondage donné) rend minimum la variance d'échantillonnage d'un paramètre privilégié. Plus généralement on peut songer à rendre minimum une certaine perte d'information.

Tant qu'on déduit celle-ci de la variance en substituant à $(x_i - x_j)^2/2$ une fonction $f(x_i, x_j, \dots)$ on est ramené aux problèmes classiques, moyennant quelque changement de paramètre; par exemple, au lieu d'effectifs de strate proportionnels aux

$$\sqrt{v_h C_h^2 / v_h - 1}$$

on les prendrait proportionnels aux $\sqrt{\gamma_h}$, dans le cas du sondage stratifié (Neyman 1934).

Nous avons montré ailleurs (T.5, Ch.3) comment à σ_h pouvaient être substitués d'autres paramètres, quand le but immédiat du sondage était d'estimer, non plus un certain ξ

$$\xi = \sum_i \sum_j x_{hj}$$

mais par exemple les σ_h eux-mêmes (en vue d'estimer ξ par un autre sondage).

2/ - Un problème nouveau.

Examinons ici le cas où l'on estime (avec biais) la plus grande valeur de X ; les cas de la plus petite valeur et de l'étendue sont analogues. La perte d'information sera définie comme ci-dessus (Chapitre IV, § II, 4) pour une strate

$$[a - E \hat{X}]_h \quad (\text{strate } h)$$

Soit $\hat{X}$ le plus grand des $\hat{X}$ de strate, c'est-à-dire la plus grande valeur de l'échantillon. La perte d'information correspondante est manifestement :

$$p = a - E \hat{X} = a - \sum \varpi_h E \hat{X}_h$$

ϖ_h étant la probabilité que le plus grand des X soit dans la strate h .

La stratification optimum serait celle qui (pour un coût $\sum_h n_h C_h$ ou un effectif $\sum_h n_h$ donné) provoquerait la perte d'information minimum. Pratiquement on connaît une variable Y en corrélation étroite avec X et on stratifie suivant des tranches de valeurs de Y . Si $\hat{X}$ a de fortes chances de se trouver dans la strate 1, on "gonflera" l'échantillon de la strate 1, mais on hésitera à priver les autres strates de tout échantillon.

3/ - La solution.

Il est possible de montrer que l'effectif optimum n_h par strate serait, en première approximation :

a) proportionnel à la racine carrée d'un certain écart moyen entre les x_{hj} ;

b) inversement proportionnel à la racine carrée du coût moyen d'enquête par unité de sondage;

c) proportionnel à la racine carrée de l'effectif v_h de la strate;

d) proportionnel à la racine carrée de la probabilité ϖ_h .

Le calcul est facile quand on suppose les n_h petits à côté des v_h , ou les n à côté des v (l'indice h étant sous-entendu) : la loi de distribution du plus grand $\hat{X}$ des X échantillon est $[F(X)]^n = G(\hat{X})$ où F prend

elle-même les valeurs $\left[0, \frac{1}{v}, \frac{2}{v} \dots \frac{v-1}{v}, 1\right]$ pour les valeurs croissantes $z_1, z_2 \dots z_{v-1}, z_v = a$ de la variable X . $G(\hat{X})$ prend donc les valeurs en palier

$$0, \left(\frac{1}{v}\right)^n, \left(\frac{2}{v}\right)^n \dots \left(\frac{v-1}{v}\right)^n, 1$$

D'où

$$a - E \hat{X} = (z_v - z_{v-1}) \left(1 - \frac{1}{v}\right)^n + (z_{v-1} - z_{v-2}) \left(1 - \frac{2}{v}\right)^n + \dots + (z_2 - z_1) \left(1 - \frac{v-1}{v}\right)^n$$

Remplaçons $\left(1 - \frac{j}{v}\right)^n$ par $e^{-\frac{nj}{v}} = u^j$ (en première approximation).

Admettons les $(z_j - z_{j-1})$ tous de l'ordre de δ ; posons $f = n/v$, $u = e^{-f}$; il vient

$$a - E \hat{X} \sim \delta \cdot (u + u^2 + \dots + u^{n-1}) \sim \delta \cdot u(1 - u)^{-1}$$

Le minimum de la perte d'information (toutes strates)

$$\Sigma_h \left[\varpi \frac{\delta}{1 - u} \right]_h$$

lié par

$$\Sigma_h [c v f]_h = c_0$$

est donné par

$$(1 - u)^2 \div u \frac{\varpi \delta}{c v} \left]_h$$

d'où approximativement $f^2 \div \frac{\varpi \delta}{c v} \left]_h$

4/ - Evaluation des ϖ_h .

En réalité les ϖ_h sont très différents les uns des autres lorsque la stratification est faite efficacement. Il serait peu réaliste de n'en pas tenir compte sous prétexte qu'on ne connaît pas leurs valeurs.

Lorsqu'on se sent capable de donner des valeurs numériques subjectives $q_1 \dots q_h$, aux probabilités a priori que $a_1 \dots a_h$ soit le plus grand des a , il résulte de la formule de Bayes que ϖ_h et q_h diffèrent peu l'un de l'autre, si la stratification est bien faite; l'emploi de coefficients arrondis 0,8; 0,5; ou 0,1 est très suffisant en pratique.

CHAPITRE VI

SONDAGES DIVERS

I - SONDAGES A DEUX PHASES -

A. 1/ - Généralités.

On dit que le sondage est à deux phases lorsque l'enquête porte sur un caractère auxiliaire $\underline{Y}$ soumis à un sondage double (1ère et 2ème phases) - et sur les caractères principaux X soumis au sondage à la seconde phase seulement de l'enquête. Ceci constitue un procédé pratique d'enquête (sociale par exemple) pourvu qu'il existe entre X et Y une forte corrélation, et que le coût des opérations d'enquête concernant Y sans X soit très faible comparé à celui concernant à la fois X et Y . Par exemple ce sera le cas lorsque Y est obtenu en consultant des documents, alors qu'il faut envoyer un enquêteur sur le terrain pour recueillir la valeur de X .

2/ - Cas où l'échantillon de la 2ème phase est extrait du grand échantillon de la 1ère phase.

Supposons que la 2ème phase porte sur un petit échantillon prélevé dans le grand échantillon tiré à la 1ère phase. Entre les deux phases on peut stratifier le grand échantillon suivant les valeurs de Y , et le petit est extrait de cette sous-population stratifiée; cette méthode est due à Neyman (1938)⁽¹⁾.

Soit $\bar{X}_2$ l'estimateur de $\bar{x}$ après les deux phases, soit $\bar{X}_1$ l'estimateur qu'on pourrait former pour $\bar{x}$ avec le grand échantillon, si l'on recueillait les données x_i correspondantes (ce qui n'est justement pas le cas). Admettons que $\bar{X}_1$ et $\bar{X}_2$ soient sans biais; il vient(2) :

$$\begin{aligned} v \bar{X}_2 &= \mathcal{E}(\bar{X}_2 - \bar{x})^2 \\ &= \mathcal{E}(\bar{X}_2 - \bar{X}_1)^2 + \mathcal{E}(\bar{X}_1 - \bar{x})^2 + 2 \mathcal{E}(\bar{X}_2 - \bar{X}_1)(\bar{X}_1 - \bar{x}) \end{aligned}$$

(1) N2. La théorie de Neyman se limite à des tirages bernoulliens.

(2) Ce calcul est emprunté à T. 4.

Mais on a :

$$\mathcal{E}[(\bar{X}_2 - \bar{X}_1) \cdot (\bar{X}_1 - \bar{x})] = 0$$

le tirage du petit échantillon ne dépendant pas des résultats de la 1ère phase.

D'autre part, pour un grand échantillon déterminé, on a :

$$E_1(\bar{X}_2 - \bar{X}_1)^2 = V_1 \bar{X}_2$$

la lettre V désignant une variance aléatoire contrairement à la lettre $\mathcal{V}$. Finalement il vient :

$$\mathcal{V} \bar{X}_2 = \mathcal{V} \bar{X}_1 + \mathcal{E}(V_1 \bar{X}_2)$$

a) $\mathcal{V} \bar{X}_1$ dépend de l'effectif du grand échantillon; par exemple si on le tire par simple tirage au sort (bernoullien ou exhaustif) $\mathcal{V} \bar{X}_1$ est fonction de la taille m dudit échantillon :

$$\mathcal{V} \bar{X}_1 = \frac{\sigma^2}{m} \quad \text{ou} \quad \frac{\sigma^2}{m} \frac{\mu - m}{\mu - 1}$$

μ étant l'effectif de la population.

b) $V_1 \bar{X}_2$ dépend des effectifs du grand et du petit échantillons; par exemple, si le grand échantillon est découpé en strates d'effectif M_h , où les tirages effectués sont exhaustifs, on a :

$$V_1 \bar{X}_2 = \sum_h \left(\frac{M_h}{m} \right)^2 \frac{M_h S_h^2}{M_h - 1} \left(\frac{1}{n_h} - \frac{1}{M_h} \right)$$

avec $\sum M_h = m$

n_h = effectif du petit échantillon, strate h .

S_h^2 = variance du grand échantillon, strate h .

On en déduit :

$$\mathcal{E}(V_1 \bar{X}_2) = \sum_h \mathcal{E} \left[\frac{M_h^2}{m^2} \left(\frac{1}{n_h} - \frac{1}{M_h} \right) \right] \mathcal{E} \left[\frac{M_h S_h^2}{M_h - 1} \right]$$

ens'appuyant sur le fait que les deux crochets ont une covariance nulle, parce que, quel que soit M_h , l'espérance mathématique du dernier crochet est : $\mu_h \sigma_h^2 / \mu_h - 1$

Sil'on désigne alors par p_h la probabilité de tirer une unité de la

strate h lors du tirage initial, on a finalement :

$$\mathcal{E}(M_h) = m p_h; \quad \mathcal{E}(M_h^2) = (m p_h)^2 + m p_h q_h \frac{\mu - m}{\mu - 1}$$

$$\mathcal{E}(V_1 \bar{X}_2) = \sum_h \frac{\mu_h \sigma_h^2}{\mu_h - 1} \left[\frac{p_h^2}{n_h} - \frac{p_h}{m} + \frac{\mu p_h q_h}{(\mu - 1)n_h} \left(\frac{1}{m} - \frac{1}{\mu} \right) \right]$$

Remarque 1.

Pour $m = \mu$, on retrouve la variance du sondage stratifié limite

$$\lim \mathcal{E} V_1 \bar{X}_2 = \mathcal{V} \bar{X}_s = \sum_h \left(\frac{\mu_h}{\mu} \right)^2 \frac{\mu_h \sigma_h^2}{\mu_h - 1} \left(\frac{1}{n_h} - \frac{1}{\mu_h} \right)$$

Remarque 2.

Ce calcul suppose d'ailleurs n_h inférieur ou au plus égal à M_h . Or si l'on se donne à l'avance m et les n_h (avec $\sum_h n_h < m$) il y a souvent une probabilité non nulle (et parfois assez grande) que n_h dépasse M_h .

Le calcul n'est donc correct que dans des cas limités, où m reste en fait voisin de μ .

Mais il faut distinguer deux possibilités :

$M_h = 0$: Cas extrême où la strate (h) ne serait pas représentée du tout dans l'échantillon; si μ_h est très petit, l'inconvénient pratique est minime; si μ_h n'est pas petit, le cas ne peut se produire qu'exceptionnellement (si m est lui-même assez grand).

$0 < M_h < n_h$: (cas qui disparaît d'ailleurs si l'on remplace les tirages exhaustifs par des tirages bernoulliens à la deuxième phase). La contribution de la strate à la variance doit être considérée comme nulle (et non pas négative). L'effectif réel de la strate en 2ème phase est ramené de n_h à M_h (économie d'argent et perte d'information). Moyennant quoi, tout rentre dans l'ordre(1).

Dans la suite, on fera tous les calculs formellement (avec des

(1) Nota - Se donner m et choisir les n_h , une fois connus les M_h , ou encore se donner les n_h et augmenter m jusqu'à ce que tous les M_h soient supérieurs aux n_h , sont aussi des procédés possibles mais défectueux car : les estimateurs employés ne sont plus sans biais; la variance voit son expression modifiée.

tirages exhaustifs aux deux phases) sans chercher à savoir à partir de quel moment on sort du domaine des choses possibles.

B - VARIANCE ET PERTE D'INFORMATION.

On désire savoir quand on a le droit d'écrire :

$$\mathcal{V} \bar{X}'_2 = \mathcal{V} \bar{X}_2 + \mathcal{E}(\bar{X}'_2 - \bar{X}_2)^2$$

en envisageant deux échantillons $(m, n_h, \bar{X}_2)$ et $(m', n'_h, \bar{X}'_2)$ avec

$$m \geq m', \quad n_h \geq n'_h$$

1/ - Etude du cas $m = m'$.

$$\text{Alors } \mathcal{V} \bar{X}'_1 = \mathcal{V} \bar{X}_1$$

$$\text{D'où : } \mathcal{V} \bar{X}'_2 - \mathcal{V} \bar{X}_2 = E(V_1 \bar{X}'_2) - E(V_1 \bar{X}_2)$$

$$\text{c'est-à-dire : } = \sum_h E \left(\frac{M_h S_h^2}{M_h - 1} \right) E \left(\frac{M_h}{m} \right) \left(\frac{1}{n_h} - \frac{1}{n_h} \right)$$

D'autre part, $\bar{X}'_2$ étant l'estimateur $\sum M_h \bar{X}'_{2h} / m$, et les (n'_h) unités étant tirées au sort parmi les (n_h) , on a :

$$E_2(\bar{X}'_2 - \bar{X}_2)^2 = \sum_h \left(\frac{M_h}{m} \right)^2 \frac{n_h s_h^2}{n_h - 1} \left(\frac{1}{n'_h} - \frac{1}{n_h} \right)$$

en désignant par (s_h^2) les variances de strates de l'échantillon (n_h) . D'où :

$$\begin{aligned} E_1 E_2 (\bar{X}'_2 - \bar{X}_2)^2 &= \sum_h E_1 \left[\left(\frac{M_h}{m} \right)^2 \frac{n_h s_h^2}{n_h - 1} \left(\frac{1}{n'_h} - \frac{1}{n_h} \right) \right] \\ &= \sum_h \left(\frac{M_h}{m} \right)^2 \frac{M_h s_h^2}{M_h - 1} \left(\frac{1}{n'_h} - \frac{1}{n_h} \right) \end{aligned}$$

Enfin :

$$E E_1 E_2 (\bar{X}'_2 - \bar{X}_2)^2 = \sum_h E \left(\frac{M_h}{m} \right)^2 E \left(\frac{M_h s_h^2}{M_h - 1} \right) \left(\frac{1}{n'_h} - \frac{1}{n_h} \right)$$

c'est-à-dire :

$$E E_1 E_2 (\bar{X}'_2 - \bar{X}_2)^2 = \mathcal{V} \bar{X}'_2 - \mathcal{V} \bar{X}_2$$

Conclusion.

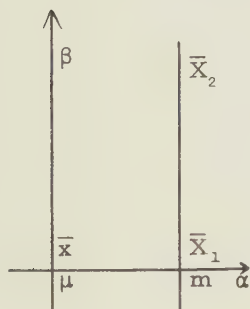
Lorsqu'on maintient constant l'effectif du grand échantillon, on a le droit de considérer la variance comme une perte d'information.

2/ - Représentation graphique.

On peut repérer $\bar{X}_2$ en coordonnées cartésiennes (α, β) avec :

$$\alpha = \overline{\mu m} = \mathcal{V}(\bar{X}_1)$$

$$\beta = \overline{m\bar{X}_2} = \mathcal{E}V_1\bar{X}_2$$



L'axe $O\alpha$ porte les points représentatifs de la première phase du sondage ; les écarts le long de cet axe sont des pertes d'information.

Pour $m = \text{Cste}$, on vient de voir que les différences $\mathcal{V}(\bar{X}_1) - \mathcal{V}(\bar{X}_2)$, c'est-à-dire les écarts, parallèlement à l'axe des β , sont également des pertes d'information (en particulier pour $m = \mu$: sondage stratifié limite) : ces verticales (y compris l'axe des β) sont des "génératrices d'information".

3/ - Recherche d'une seconde famille de génératrices
($n_h = \text{constantes}$)

On va maintenir les n_h constants et faire varier m (à partir de $m = \mu$, qui correspond à $\bar{X}_2 = \bar{X}_s$).

On a :

$$\mathcal{V}\bar{X}_1 = \frac{\mu}{\mu - 1} \left(\frac{1}{m} - \frac{1}{\mu} \right)$$

avec :

$$\sigma^2 = \sum_h p_h \sigma_h^2 + \sum_h p_h (\bar{x}_h - \bar{x})^2$$

D'où :

$$\begin{aligned} \mathcal{V}\bar{X}_2 - \mathcal{V}\bar{X}_s &= \left[\frac{\mu}{\mu - 1} - \sum_h p_h \frac{\mu_h \sigma_h^2}{\mu_h - 1} + \sum_h p_h \frac{\mu_h \sigma_h^2}{\mu_h - 1} \frac{\mu}{\mu - 1} \frac{1}{n_h} \right] \left(\frac{1}{m} - \frac{1}{\mu} \right) \\ &= \frac{\mu}{\mu - 1} \left[\sum_h p_h (\bar{x}_h - \bar{x})^2 + \frac{\mu_h}{\mu_h - 1} p_h q_h \sigma_h^2 \left(\frac{1}{n_h} - \frac{1}{\mu_h} \right) \right] \left(\frac{1}{m} - \frac{1}{\mu} \right) \end{aligned}$$

4/ - Conséquences.

Lorsqu'on figure un sondage à deux phases par un grillage (α, β) et qu'on porte les longueurs $\alpha = \mathcal{V}\bar{X}_1$, $\beta = \mathcal{V}\bar{X}_s$ sur les deux axes, on n'a ni $\mathcal{V}\bar{X}_2 = \alpha + \beta$, ni $\alpha + \beta + \gamma$, mais :

$$\mathcal{V} \bar{X}_2 = \alpha + \beta - \gamma$$

avec

$$\gamma = \frac{\mu}{\mu - 1} \sum_h p_h \sigma_h^2 \left[1 - \frac{\mu_h}{\mu_h - 1} \left(\frac{1}{n_h} - \frac{1}{\mu_h} \right) \right] \left(\frac{1}{m} - \frac{1}{\mu} \right)$$

a) Il arrive que les p_h soient tous égaux, soit $p_h = p$; auquel cas on a :

$$\beta = p \sum p \sigma_h^2 \frac{\mu_h}{\mu_h - 1} \left(\frac{1}{n_h} - \frac{1}{\mu_h} \right)$$

d'où

$$\sigma^2 \gamma = \alpha \left[\sum_h p \sigma_h^2 - \frac{\beta}{p} \right]$$

équation d'un paraboloïde hyperbolique dans l'espace euclidien (α, β, γ) . Avec $\alpha = \beta = 0$, on retrouve l'axe des β (sondages stratifiés limites). En revanche, si $n_h = \mu_h$ quel que soit h , il vient : $\gamma = \alpha (\sum p \sigma_h^2 / \sigma^2)$ qui se projette sur l'axe $O\alpha$ mais n'est pas cet axe).

[D'où :

$$\mathcal{V} \bar{X}_2 = \alpha + \beta - \gamma = \alpha [\sum p (\bar{x}_h - \bar{x})^2 / \sigma^2]$$

alors qu'on s'attendait à $\mathcal{V} \bar{X}_2 = \mathcal{V} \bar{X}_1 = \alpha$. En réalité, on a : $\sum n_h = \mu$, donc $m = \mu$, $M_h = \mu_h$, donc $\alpha = 0$].

Ainsi, si $p_h = p$, il existe une image métrique euclidienne du grilage du sondage à deux phases, analogue à celle du sondage à deux degrés, mais l'information (perdue) est $\alpha + \beta - \gamma$, au lieu de $\alpha + \beta + \gamma$.

b) En général les p_h ne sont pas égaux; et l'équation en α, β, γ n'existe plus.

On retrouve le même phénomène que pour le sondage à deux degrés (où en dehors du cas où $v_h = \bar{v}$, l'image euclidienne demande des conventions adéquates).

On peut toujours poser :

$$\beta = \sum_h \beta_h, \quad \beta' = \sum_h \beta_h / p_h, \quad b = \sum_h p_h \sigma_h^2$$

avec :

$$\beta_h = p_h^2 \frac{\mu_h \sigma_h^2}{\mu_h - 1} \left(\frac{1}{n_h} - \frac{1}{\mu_h} \right)$$

$$\text{d'où} \quad \sigma^2 \gamma = \alpha (b - \beta')$$

équation d'un paraboloïde hyperbolique; les génératrices ($\alpha = \text{constantes}$) correspondent à m constant; celles ($\beta' = \text{constantes}$) correspondent aux n_h constants. Mais la perte d'information symétrique

$$\mathcal{V} \bar{X} \quad \text{est} \quad \alpha + \beta - \gamma \quad \text{et non} \quad \alpha + \beta' - \gamma$$

Avec $L (= 1/p)$ strates, il faut utiliser $(L+2)$ dimensions ($\alpha \ \beta_h \ \gamma$) et une surface d'équation

$$\sigma^2 \gamma = \alpha (b - \sum \beta_h / p_h)$$

$$\text{d'où la perte} \quad \mathcal{V} \bar{X}_2 = \alpha + \sum \beta_h - \gamma$$

On voit facilement que $\mathcal{E}(\bar{X}_s - \bar{x})(\bar{X}_2 - \bar{X}_s) > 0$

$$\mathcal{E}(\bar{X}_1 - \bar{x})(\bar{X}_2 - \bar{X}_1) > 0$$

Peut-on parler du gain d'information dû à une stratification complète ?

Lorsqu'on peut stratifier toute la population, on ne fait pas de sondage à deux phases : au sondage stratifié d'effectifs n_h correspond la perte d'information $\mathcal{V} \bar{X}_s$. On vérifierait que $\mathcal{V} \bar{X}_2 - \mathcal{V} \bar{X}_s = \alpha - \gamma$ est positif. On peut considérer que cette différence mesure l'information gagnée à stratifier, ou plutôt la perte d'information (à partir de $\bar{X}_s$) due à une stratification ne portant que sur m unités (au lieu de μ).

Ce type de question sera examiné au Chapitre VII, IIème partie.

II - INDICATIONS SUR QUELQUES AUTRES PLANS -

On a signalé au Ch. III l'intérêt qu'il pouvait y avoir parfois à figurer une famille de plans de sondage sur un treillis en quelque sorte pléthorique; d'autre part on connaît la forme générale des pertes d'information (sous certaines hypothèses restrictives, Ch. IV); disons un mot de la variance d'échantillonnage d'estimateurs sans biais relatifs à des méthodes de sondage (assez courantes) sortant plus ou moins du cadre du chapitre V.

1/ - Sondages stratifiés "optimum".

Soit une suite de plans de sondage stratifiés dont chacun est "optimum" au sens de Bowley, Neyman ou Yates, l'effectif ou le coût décroissant régulièrement; on peut représenter cette suite en portant $\mathcal{V} \bar{X}$ sur un axe (voir Ch. V, fin du § II).

On peut également repérer les points ($\text{opt } n_h$) sur le treillis des sondages stratifiés correspondants, sommets d'une "trajectoire" en ligne brisée qui n'est pas un axe d'information (voir Ch. III, § II. 5.b).

$\forall \bar{X}$ est combinaison linéaire des pertes par strate et vérifie la relation de récurrence.

2/ - Stratification a posteriori.

Sur le treillis précédent on envisage des "surfaces d'onde" ($\Sigma n_h = n$); sur chacune d'elles on donne une distribution de probabilité pour l'estimateur $\bar{X}_s$ et la variance $V\bar{X}_s$; la taille n de l'échantillon étant donnée, les n_h sont aléatoires (tirage au sort de l'échantillon dans une urne où toutes les strates sont mélangées). On pose $\mathcal{E} V\bar{X}_s = p(\bar{X})$.

Si certains des n_h sont nuls ($\bar{X}_h$ indéterminés), on convient à l'avance de l'expression à donner à $\bar{X}_s$ pour qu'il ne soit pas indéterminé (sans introduire de biais).

Ici encore il est clair que $p(\bar{X})$ est combinaison linéaire de pertes par strate $\mathcal{E}(V\bar{X}_h)$ et vérifie la relation de récurrence.

CHAPITRE VII

LES INFORMATIONS SUPPLÉMENTAIRES

I - LE CHANGEMENT D'ESTIMATEUR -

A - GENERALITES.

1/ - Rappel de la technique classique.

On sait que, pour un échantillon d'ores et déjà tiré, le choix d'un estimateur plus efficace que l'estimateur courant (construit avec le seul plan d'échantillonnage : probabilités de tirage, effectifs, etc.) est rendu parfois possible grâce à des "informations supplémentaires". Les cas habituels sont :

- l'estimation par ratio (ou par quotient);
- l'estimation par une formule de régression;
- l'estimation par une formule de stratification (a posteriori).

Ces dernières années Žarković a ajouté à cette liste l'estimation par différence (Réf. Z1).

Dans tous ces cas, on se propose d'estimer pour une population donnée une certaine moyenne $\bar{x}$ d'une variable X, alors qu'on a des informations sur une variable Y en corrélation étroite avec X. Rien n'empêche de substituer à $\bar{x}$ et X des expressions ζ et Z plus générales.

Même lorsqu'il est question de strates, elles n'interviennent qu'après coup. Aussi peut-on supposer (pour simplifier) que le sondage est à un seul degré, sans strate, avec probabilités égales (treillis réduit à un axe, commun à tous estimateurs).

On connaît, non seulement les x_i des unités-échantillon, mais aussi leurs y_i , ainsi que la moyenne $\bar{y}$ pour la population entière, et plus généralement toute la distribution des y_i .

On indiquera d'autres cas où des renseignements d'une nature différente permettent de réduire la variance d'échantillonnage en changeant d'estimateur.

On peut rapprocher enfin cette question de celle de l'estimation du maximum de vraisemblance et plus généralement du recours à des hypothèses sur la loi mathématique qui réglerait la distribution des x_i , hypothèses permettant de changer d'estimateur et (par là) de réduire la variance.

2/ - Problème.

A-t-on le droit d'appeler gain d'information la réduction de variance ? Autrement dit : s'agit-il bien d'une quantité d'information ?

Si Z et Z' sont deux estimateurs sans biais de ζ , avec $\varphi Z > \varphi Z'$, il est exact que $\varphi Z - \varphi Z'$ est une perte d'information générale (au sens des ch. II et III). Mais il n'est pas en général exact que ce soit égal à $\mathcal{E}(Z - Z')^2$.

Or, il importe assez peu que $(\varphi Z - \varphi Z')$ soit une information dans l'optique d'une suite de plans de sondage d'une famille F , alors qu'on a changé d'optique et qu'on n'envisage qu'un seul plan (et même un seul échantillon). Ce n'est plus en augmentant la taille de l'échantillon qu'on passe du point Z au point Z' . Il n'en serait pas moins intéressant de pouvoir représenter ZZ' et ζ par trois points alignés, $\varphi Z - \varphi Z'$ étant la distance (ZZ') , mesurée avec la même métrique que (ζZ) ou (ZZ') , c'est-à-dire $\mathcal{E}(Z - Z')^2$.

On dirait alors que $(\varphi Z - \varphi Z')$ est l'information supplémentaire.

Dans les divers cas étudiés s'est retrouvé le résultat que voici.

THEOREME -

Si l'estimateur Z' est substitué à Z , l'un et l'autre sans biais, avec $\varphi Z' < \varphi Z$, $\varphi Z - \varphi Z'$ n'est pas en général une information supplémentaire, à moins qu'on n'ait :

$$\mathcal{E}(Z' - \zeta)(Z - Z') = 0 ;$$

En effet, si l'on convient de poser, comme définition de l'information supplémentaire :

$$\varphi Z - \varphi Z' = \mathcal{E}(Z - Z')^2 ,$$

c'est cette relation qui exprime l'orthogonalité des deux vecteurs aléatoires $\overline{\zeta Z'}$ et $\overline{Z'Z}$.

Suivant les problèmes traités, Z' est bien déterminé ou au contraire dépend de paramètres, par exemple linéairement (auquel cas on considèrera un point Z' décrivant une multiplicité linéaire). Nous reviendrons plus loin sur le cas où Z' est biaisé.

3/ - Exemple.

Supposons tiré un échantillon d'effectifs n_h dans chaque strate (notations du Ch. V). Parmi tous les estimateurs linéaires sans biais $Z_\lambda = \sum \lambda_h \bar{x}_h$ dépendant des λ_h liés par $\sum_h \lambda_h \bar{x}_h = \bar{x}$, l'estimateur de variance minimum Z° serait tel que $d[\sum \lambda_h \bar{x}_h + \mu \sum \lambda_h \bar{x}_h] = 0$ (μ de Lagrange) ce qui signifie que les λ_h devraient être proportionnels aux $n_h \bar{x}_h / \tau_h^2$, avec $\tau_h^2 = v_h \sigma_h^2 / v_h - 1$.

Les informations supplémentaires à connaître seraient ici les $\tau_h^2 / \bar{x}_h$ ou (si tirages bernoulliens) les $\sigma_h^2 / \bar{x}_h$.

Pratiquement (depuis Bowley) on pose $\lambda_h = v_h / v$ pour éliminer le biais avec des poids indépendants de la variable X étudiée, en l'absence d'informations supplémentaires; posons $Z = \sum v_h \bar{x}_h / v$;

$$\sum Z - \sum Z^\circ \quad (\geq 0)$$

représente l'information perdue faute de connaître les $\tau_h^2 / \bar{x}_h$; alors que

$$\sum Z_\lambda - \sum Z \quad (\geq 0)$$

ne devrait pas être tenu comme un gain ou une perte d'information faute d'user des poids (v_h / v) .

Remarque.

Supposons les n_h choisis (à la Neyman) proportionnellement aux $(\tau_h v_h / \sqrt{C_h})$ d'une certaine variable X . L'estimateur de Neyman correspond à $\lambda_h = v_h / v$; alors que l'estimateur de variance minimum correspond à λ_h proportionnel à

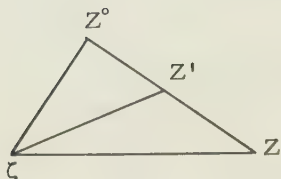
$$v_h \bar{x}_h / \tau_h \sqrt{C_h}$$

la coïncidence exigeant que le coût moyen d'enquête par strate C_h soit inversement proportionnel à $(v_h \gamma_h^2 / v_h - 1)$, en gros à γ_h^2 , carré du coefficient de variation par strate.

Ceci est à noter, d'autant plus que l'optimum au sens de Neyman ne concerne jamais qu'une variable particulière X ; or cet "optimum" minimise la variance d'un estimateur très particulier: celui qui reste sans biais pour toutes variables X ; après quoi un estimateur Z° devrait être substitué à Z .

4/ - Retour au cas général.1ère interprétation.

Dans l'espace euclidien auxiliaire où $\overline{\zeta Z}^2$ représente $\mathcal{V} Z$, figurons par Z' l'estimateur de ζ obtenu grâce aux informations supplémentaires.



Supposons $\mathcal{V} Z' < \mathcal{V} Z$

mais $\mathcal{V} Z - \mathcal{V} Z' \neq \mathcal{E}(Z' - Z)^2$

Soit Z^0 la projection de ζ sur ZZ' , c'est-à-dire $\mathcal{V} Z^0$ le minimum de $\mathcal{V} Z'$.

$$\mathcal{V} Z' = \mathcal{E}(Z^0 - \zeta)^2 + \mathcal{E}(Z' - Z^0)^2$$

$$\mathcal{V} Z = \mathcal{E}(Z^0 - \zeta)^2 + \mathcal{E}(Z - Z^0)^2$$

Par analogie avec la relation de récurrence de l'information d'après Schutzenberger, on dira que :

$$\mathcal{E}(Z' - Z^0)^2 \text{ ou } \mathcal{E}(Z - Z^0)^2$$

est une perte d'information par rapport à l'estimateur Z^0 .

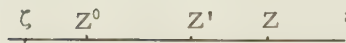
En revanche,

$$\mathcal{V} Z - \mathcal{V} Z' = \mathcal{E}(Z - Z^0)^2 - \mathcal{E}(Z' - Z^0)^2$$

ne sera pas considéré comme une information supplémentaire.

2ème interprétation.

Dans l'espace euclidien auxiliaire où ζZ représente $\mathcal{V} Z$, les points $\zeta Z^0 Z' Z$ sont alignés. La distance $Z'Z$ est mesurée par $\mathcal{V} Z - \mathcal{V} Z'$ mais non par $\mathcal{E}(Z - Z')^2$, à moins que Z' ne coïncide avec Z^0 .



On ne considèrera donc pas la droite $\zeta Z^0 Z' Z$ comme un véritable axe d'information⁽¹⁾.

(1) Construire un triangle $\zeta ZZ'$ avec $\zeta Z = \mathcal{V} Z$, $\zeta Z' = \mathcal{V} Z'$, $ZZ' = \mathcal{E}(Z - Z')^2$ n'est possible que si le coefficient de corrélation entre Z et Z' est compris entre 0 et une limite supérieure au plus égale à 1 (cas $\mathcal{V} Z = \mathcal{V} Z'$). (espace euclidien).

B - ETUDE DE CAS OU LES DEUX ESTIMATEURS SONT SANS BIAIS.

On retrouve le triangle rectangle $\zeta Z^0 Z$ ou plutôt $\overline{X}X^0\overline{X}$ dans les cas suivants :

1/ - Estimation par différence.

L'estimateur par différence (de Žarković) X est tel que :

$$X^0 - \overline{X} = \overline{y} - \overline{Y}$$

Il n'y a ici orthogonalité que dans un cas particulier :

$$\rho_{xy} = \rho = \sigma_y / \sigma_x$$

bien que la variance soit réduite dès que :

$$\rho > \sigma_y / 2\sigma_x$$

Il suffit d'écrire la relation bien connue :

$$\rho \sigma_x = b \sigma_y$$

où b est la pente de la droite de régression de x en y , pour voir que la structure particulière pour laquelle il y a gain d'information au sens strict est celle où

$$\underline{b = 1}$$

2/ - Estimation par ratio : $X^0 = \overline{X} \overline{y} / \overline{x}$

On supposera d'emblée que le biais est négligeable (si non nul), c'est-à-dire :

$$\rho \gamma_x - \gamma_y = 0$$

c'est-à-dire

$$b = \overline{x} / \overline{y}$$

qui signifie que la droite de régression de X en y passe par l'origine. C'est le cas où la variance $\mathcal{V} X^0$ diffère très peu de :

$$\mathcal{V} \overline{X} \left[1 - \rho^2 + \left(\frac{\gamma_y}{\gamma_x} - \rho \right)^2 \right] = \mathcal{V} \overline{X} (1 - \rho^2)$$

(égalité rigoureuse si la distribution XY a la structure du N°3).

3/ - Estimation par régression : $X^0 = \bar{X} + b(\bar{y} - \bar{Y})$.

On s'en tient au cas "idéal" où b lui-même est connu (et non une estimation B sur échantillon) tout en postulant que la distribution à deux dimensions (x_i, y_i) a la structure particulière suivante :

- la distribution des y_i est quelconque;
- la droite de régression de X en y est quelconque : $x = a + by$;
- on a $X = a + by_i + \alpha\theta$ avec $\mathcal{E} \theta = 0$; $\mathcal{E} \theta^2 = 1$;
- la distribution de θ est indépendante de celle de Y ;
- σ_i dépend de y_i .

On a alors à nouveau $\mathcal{V} X^0 = \mathcal{V} \bar{X} (1 - \rho^2)$.

4/ Sondage bernoullien avec identification.

On désignera ainsi le cas où les boules portent un signe distinctif (un numéro d'identité) permettant de reconnaître, parmi n boules tirées, celles qui ont été tirées 2 fois, 3 fois, ...

Alors on a intérêt à choisir comme estimateur, au lieu de la moyenne X_n , le rapport suivant :

$$X_n^0 = \frac{\text{total des } x \text{ tirés (chacun compté une fois)}}{\text{nombre de boules distinctes tirées}}$$

Cet estimateur, qui s'identifie (avec des probabilités diverses) à ceux des sondages exhaustifs d'effectif n , $n-1$, $n-2$, ... est évidemment sans biais; et sa variance est une combinaison linéaire des variances desdits estimateurs exhaustifs, soit :

$$V_n, \quad V_{n-1}, \quad V_{n-2}, \dots$$

On trouve ainsi :

$$\mathcal{V} X_2^0 = \frac{1}{v} V_1 + \frac{v-1}{v} V_2$$

$$\mathcal{V} X_3^0 = \frac{1}{v^2} V_1 + 3 \frac{v-1}{v^2} V_2 + \frac{(v-1)(v-2)}{v^2} V_3$$

$$\mathcal{V} X_4^0 = \frac{1}{v^3} V_1 + 7 \frac{v-1}{v^3} V_2 + 6 \frac{(v-1)(v-2)}{v^3} V_3 + \frac{(v-1)(v-2)(v-3)}{v^3} V_4$$

$$\mathcal{V} X_5^0 = \frac{1}{v^4} V_1 + 15 \frac{v-1}{v^4} V_2 + 25 \frac{(v-1)(v-2)}{v^4} V_3 + 10 \frac{(v-1) \dots}{v^4} V_4 + \frac{(v-1) \dots}{v^4} V_5$$

Le coefficient de V_i sur la ligne $\mathcal{V}X_n^0$ est la probabilité de tirer i boules distinctes en n tirages. D'où (en explicitant) :

$$\mathcal{V}X_2^0 = \frac{\sigma^2}{2} = \mathcal{V}\bar{X}_2$$

$$\mathcal{V}X_3^0 = \frac{\sigma^2}{3} \frac{2v-1}{2v} = \frac{\sigma^2}{3} - \frac{\sigma^2}{6v} < \mathcal{V}\bar{X}_3$$

$$\mathcal{V}X_4^0 = \frac{\sigma^2}{4} \frac{v(v-1)}{v^2} = \frac{\sigma^2}{4} - \frac{\sigma^2}{4v} < \mathcal{V}\bar{X}_4$$

$$\mathcal{V}X_5^0 = \frac{\sigma^2}{5} - \frac{\sigma^2}{5} \frac{9v^2 - v - 1}{6v^3} < \mathcal{V}\bar{X}_5$$

et ainsi de suite indéfiniment. La formule générale est :

$$\mathcal{V}X_n^0 = \frac{1}{v^{n-1}} [V_1 + a_2(v-1)V_2 + a_3(v-1)(v-2)V_3 + \dots + (v-1) \dots (v-n+1)V_n]$$

avec : $a_2 = 2^{n-1} - 1$

$$a_3 = (3^{n-1} - 2^n + 1)/2!$$

$$a_4 = (4^{n-1} - 3^n + 3 \cdot 2^{n-1} - 1)/3! \quad \text{etc.}$$

Le fait que les boules portent un signe d'identité apporte bien une information, c'est-à-dire que la différence des deux variances

$$\mathcal{V}\bar{X}_n - \mathcal{V}X_n^0$$

est égale à $\mathcal{E}(\bar{X}_n - X_n^0)^2$, autrement dit : $\bar{X}X_n^0$ et $\bar{X}_nX_n^0$ sont orthogonaux. En effet :

Raisonnons dans le cas où $n = 4$. Supposons qu'on ait tiré $(a \ b \ c \ d)$, on a $\bar{X} = X^0$. Si on a tiré au contraire $(a \ c \ c \ d)$, on a $X^0 = \frac{a+c+d}{3}$. Mais à cet X^0 , correspondent trois $\bar{X}$ distincts, équiprobables :

$$\frac{2a+c+d}{4},$$

$$\frac{a+2c+d}{4},$$

$$\frac{a+c+2d}{4}$$

On a évidemment :

$$\mathcal{E}(\bar{X} \cdot | X^0) = \frac{1}{3} \frac{4a+4c+4d}{4} = X^0$$

d'où suit l'orthogonalité, et le gain d'information.

X^0 est la projection de $\bar{x}$ sur la multiplicité linéaire

$$\lambda x_i + \mu x_j + (1 - \lambda - \mu) \bar{X}$$

en désignant par x_i , x_j , les x des boules tirées plusieurs fois.

5/ - Estimation du carré de l'écart-type (avec ou sans la vraie moyenne).

La question suivante nous a été posée (1956) par M. Fonsagrive, pour les sondages bernoulliens.

On possède un échantillon, avec lequel on estime le σ^2 de l'urne-mère. Quelle information m'apporte-t-on si l'on m'apprend par ailleurs la vraie valeur de la moyenne de la population ? (plus correctement : de combien est réduite la perte d'information ?)

a) Si $\bar{x} = \mu_1$ est connu, on a un échantillon de n valeurs

$$z_i = (x_i - \mu_1)^2$$

dont la moyenne $\bar{Z} = S_1(x_i - \mu_1)^2/n$

estime sans biais $\zeta = \sigma^2$, avec la variance (où σ_z^2 désigne $\mu_4^0 - \sigma^4$)

$$\mathcal{V} \bar{Z} = \begin{cases} \sigma_z^2/n & \text{tirages avec remise} \\ \frac{\sigma_z^2}{n} \frac{v-n}{v-1} & \text{tirages sans remise} \end{cases}$$

b) Si $\bar{x} = \mu_1$ n'est pas connu, mais estimé par $\bar{X}$, on estime σ^2 par

$$\frac{v-1}{v} \frac{n S^2}{n-1} = \frac{v-1}{v} \bar{Y}$$

1) Faisons tendre v vers l'infini; on sait, cf. Ch. I, § III 8 a (cas du sondage bernoullien) que :

$$\mathcal{V} \bar{Y} = \mathcal{V} \bar{Z} + 2\sigma^4/n(n-1)$$

Donc $2\sigma^4/n(n-1)$ est la réponse à la question posée; sous réserve qu'il soit bien justifié d'appeler "information" cette réduction de variance.

2) Si v est fini, on sait (cf. Ch. I, § III 8 b) que $\mathcal{V}[(v-1)\bar{Y}/v]$ est de la forme :

$$A. \mathcal{V} \bar{Z} + B \sigma^4$$

et on montre facilement que $A > 1$, $B > 0$ (dès que $v > 2$, $n \geq 2$), de sorte que la connaissance de μ_1 réduit la variance de :

$$(A - 1) v \bar{Z} + B \sigma^4$$

sans qu'on puisse considérer comme pleinement justifié l'emploi du mot "information" pour désigner cette expression.

1) Dans le cas des tirages avec remise (ou si v est infini) on va vérifier qu'on a bien :

$$v \bar{Y} - v \bar{Z} \equiv \mathcal{E} (\bar{Y} - \bar{Z})^2$$

En effet il suffit d'établir que

$$v \bar{Z} \equiv \text{Cov } \bar{Y} \bar{Z}$$

ou encore

$$\mathcal{E} \bar{Z}^2 \equiv \mathcal{E} \bar{Y} \bar{Z}$$

On a :

$$\mathcal{E} \bar{Z}^2 = \mathcal{E} \left(\frac{S z_i^2}{n} \right)^2$$

$$\bar{Y} = \frac{n}{n-1} \frac{S(z_i - z_j)^2}{n^2}$$

$$\equiv \frac{S z_i^2}{n} - \frac{2 S z_i z_j}{n(n-1)}$$

car le développement de $S(z_i - z_j)^2$ comprend $n(n-1)$ fois $(z_i z_j)$; donc dans $\bar{Y}$ figure une fois seulement le terme $(z_i z_j)$. Il vient :

$$\mathcal{E} \bar{Y} \bar{Z} = \mathcal{E} \left(\frac{S z_i^2}{n} \right)^2 - 2 \mathcal{E} \left(\frac{S z_i^2 \cdot S z_i z_j}{n^2(n-1)} \right);$$

et l'espérance du second terme est nulle puisqu'il est de la forme $\lambda \mathcal{E}(z_i^3 z_j)$, avec :

$$\mathcal{E} z_i^3 z_j = \mathcal{E} z_i^3 \mathcal{E} z_j, \quad \mathcal{E} z_j = 0 \quad \text{c. q. f. d.}$$

2) Mais dans le cas des tirages sans remise, on montrera qu'il n'en est plus ainsi, sur un simple exemple :

$$n = 2, \quad v = 3, \quad z_i = a, b, c.$$

Il vient :

$$\mathcal{E} \bar{Z}^2 = \left(\frac{a^2 + b^2}{2} \right)^2 + \left(\frac{b^2 + c^2}{2} \right)^2 + \left(\frac{c^2 + a^2}{2} \right)^2$$

D'autre part, en substituant $v\bar{Y}/v-1$ à $\bar{Y}$ pour n'avoir pas de biais, on a :

$$\begin{aligned}\frac{v}{v-1} \mathcal{E} \bar{Y} \bar{Z} &= \frac{3}{2} \left[\frac{a^2 + b^2}{2} \cdot \frac{(a-b)^2}{2} + \dots \right] \\ &= \left(\frac{a^2 + b^2}{2} \right)^2 + \dots + \frac{1}{8} (a^4 - 6a^3b + 2a^2b^2 - 6ab^3 + b^4) + \dots\end{aligned}$$

Et il n'y a aucune raison pour que les seconds termes, soit

$$[(a-b)^4 - 2(a+b)^2ab]$$

s'annulent.

Conclusion.

La connaissance du premier moment apporte un gain d'information si les tirages sont bernoulliens, une réduction de variance s'ils sont exhaustifs. Le vocable "information" apparaît finalement comme assez dangereux par les interprétations abusives qu'on serait tenté de lui donner dans les comparaisons entre estimateurs. Il n'est pas toujours facile de discerner une famille d'estimateurs Z' dont Z fait partie et dont Z° est le point le plus proche de ζ ; elle nous échappe dans les cas n°4 et 5 ci-dessus.

C - COMPARAISON ENTRE ESTIMATEURS EVENTUELLEMENT BIAISES.

1/ - Généralités et Rappels.

Tant que le biais était nul (§ A et B), variance et perte d'information étaient pris l'un pour l'autre; alors qu'en fait tout multiple de la variance est perte d'information symétrique, sans oublier l'existence des pertes d'information asymétriques; ces points prennent quelque importance pour les estimateurs biaisés.

On a vu (Ch. IV, § II) que la notion de perte d'information ne s'étendait qu'aux estimateurs qu'on peut appeler uniformément biaisés; et on sait calculer les pertes d'informations correspondantes dans des cas suffisamment étendus en pratique, notamment le cas dit isomorphe.

On a rejeté par conséquent, comme insuffisamment fondée, la technique qui consiste (pour estimer ζ) à comparer la variance $\mathcal{V}Z'$ de l'estimateur Z' sans biais et la somme $(\mathcal{V}Z + b^2)$ de l'estimateur Z affecté du biais b . Peut-être cette technique avait-elle son origine dans la théorie classique de l'estimation où (dans le cas régulier) on démontre que

$$\mathcal{V} Z + b^2 = \mathcal{E}(Z - \zeta)^2$$

est supérieur (au plus égal) à

$$\left(1 + \frac{d b}{d \zeta}\right)^2 / n H$$

théorème que, pour $b = 0$, se réduit à l'inégalité bien connue

$$\mathcal{V} Z \geq 1/n H.$$

Parmi les pertes $p(Z)$ (où ni $\mathcal{V} Z$, ni $\mathcal{V} Z + b^2$ ne figurent), on a besoin à présent de savoir en choisir une qu'on puisse le plus valablement comparer à $\mathcal{V} Z'$; soit $\bar{p}$.

Il n'est pas évident d'abord que p doive être symétrique. La technique des sondages nous conseille même, pour l'estimation d'un ratio, de faire jouer un rôle privilégié à une expression asymétrique (Ch. IV, § II. 8).

On peut tenir pour sensé le principe que, si Z se rapproche infiniment de ζ (soit $n \rightarrow \nu$, soit $n \rightarrow \infty$) $\bar{p}(Z)$ devrait être celle des $p(Z)$ qui a le contact le plus élevé avec $\mathcal{V} Z$ (principe mis en œuvre pour le ratio et la corrélation). Nous traiterons seulement deux exemples.

2/ - L'estimation par le ratio.

Par rapport aux notations du § B. 2, on a :

$$\begin{aligned} \zeta &= \bar{x}, & Z' &= \bar{X}, & Z &= \bar{X} \bar{y} / \bar{Y} \\ \bar{p}(Z) &= \bar{p} \left(\bar{y} \frac{\bar{X}}{\bar{Y}} \right) \end{aligned}$$

Il est évident (mais utile à rappeler) que

$$\bar{p}(k Z) = k^2 \bar{p}(Z).$$

Il vient donc

$$\begin{aligned} \bar{p}(Z) &= \bar{y}^2 \bar{p} \left(\frac{\bar{X}}{\bar{Y}} \right) \\ &= \bar{y}^2 \left(\frac{\bar{x}}{\bar{y}} \right)^2 \mathcal{E} \left(\frac{\bar{X}}{\bar{x}} - \frac{\bar{Y}}{\bar{y}} \right)^2 \end{aligned}$$

(cf. Ch. IV, § II. 8).

ou en posant $\sigma^2 = \mathcal{V} \bar{X}$, $\sigma'^2 = \mathcal{V} \bar{Y}$

$$\rho \sigma \sigma' = \text{Cov. } \bar{X} \bar{Y}, \quad \gamma = \sigma / \bar{x}, \quad \gamma' = \sigma' / \bar{y} :$$

$$\bar{p}(Z) = \mathcal{V} Z' - \bar{x}^2 2 \gamma \gamma' \left[\rho - \frac{\gamma'}{2 \gamma} \right]$$

D'où : $\bar{p}(Z) \leq \mathcal{V} Z'$ si $\rho \geq \gamma' / 2 \gamma$

On retrouve en toute rigueur le résultat approximatif classique avec les variances.

De même, on a :

$$\bar{p}(Z) = \mathcal{V} Z' (1 - \rho^2) + \sigma'^2 \left(\frac{\rho \sigma}{\sigma'} - \frac{\bar{x}}{\bar{y}} \right)^2 \leq \mathcal{V} Z' (1 - \rho^2)$$

Le minimum de la perte d'information est atteint quand le rayon vecteur du centre de gravité coïncide avec la droite de régression de X en y ($\rho \sigma / \sigma' = \bar{x} / \bar{y}$).

3/ - L'estimation par régression.

Si les axes sont translatés sans modifier la pente de la droite d'estimation (passant par (X, Y) , il est clair que la perte $\bar{p}(Z)$ ne doit pas se modifier (idée que nous devons à M. Fonsagrive). Donc si la droite d'estimation a pour pente

$$(\bar{x} / \bar{y} = \rho \sigma / \sigma'),$$

la perte demeure : $\bar{p}(Z) = \mathcal{V} Z' (1 - \rho^2)$

Et si cette pente est $\bar{x} / \bar{y} = t$, la perte demeure

$$\bar{p}(Z) = \mathcal{V} Z' (1 - \rho^2) + \sigma'^2 \left(t - \rho \frac{\sigma}{\sigma'} \right)^2$$

Supposons cette pente telle-même aléatoire (soit T), déterminée par l'échantillon tiré; alors $\bar{p}(Z)$ est aléatoire. C'est notamment le cas si t est l'estimation (biaisée) de b par les moindres carrés

$$T = \frac{R S}{S'} = \frac{R S S'}{S'^2}$$

Alors $\bar{p}(Z)$ n'est plus (à proprement parler) perte d'information, non plus que son espérance. - Mais l'espérance relative à l'estimation isomorphe de t par T est calculable (Ch. IV, II, § 8) :

$$\log t = \log (\rho \sigma \sigma') - \log \sigma'^2$$

$$\frac{\Delta t}{t} = \frac{\Delta (\rho \sigma \sigma')}{\rho \sigma \sigma'} - \frac{\Delta \sigma'^2}{\sigma'^2}$$

$$\bar{p}(T) = t^2 \mathcal{E} \left[\frac{R S S'}{\rho \sigma \sigma'} - \frac{S'^2}{\sigma'^2} \right]^2$$

Alors on posera :

$$\bar{p}(Z) = \vartheta Z'(1 - \rho^2) + \sigma'^2 \bar{p}(T)$$

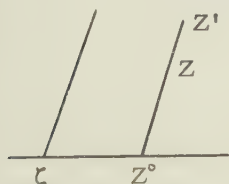
Les deux cas traités (Ratio, Régression) semblent suffire à donner une idée des simplifications que la méthode de la perte d'information introduit par rapport aux calculs classiques de variances (sans nécessiter ni approximations, ni hypothèses spéciales sur la structure des distributions).

4/ - Gain d'information est-il correct ?

On retrouve le problème discuté déjà en A et B sous une forme généralisée : $p(Z)$ et $p(Z')$ étant finalement des distances (ζZ) et ($\zeta Z'$), peut-on interpréter $|p(Z) - p(Z')|$ comme une distance ($Z Z'$) ? Si les treillis de Z et Z' sont réduits chacun à un axe, peut-on les représenter simultanément par deux axes de même origine ζ superposés ou sécants ?

Reprenons les cas (B. 2) de l'estimation par ratio et (B. 3) de la régression; le paramètre t permet de décrire la multiplicité linéaire dont font partie $\bar{X}$ et $\bar{X} \bar{y}/\bar{Y}$ et où $\bar{x}$ se projette en X° ; $1 - \rho^2 = \sin^2 \vartheta$; ϑ est l'angle de $\bar{x} \bar{X}$ et $\bar{x} X^\circ$ (première interprétation : les pertes sont les carrés des distances euclidiennes).

Ainsi, quand les pertes sont des distances euclidiennes (2ème interprétation) on ne doit pas superposer les trois axes (Z) (Z') (Z°) de crainte d'interprétations abusives. D'ailleurs il n'y a plus de raison de vouloir considérer la distance ($Z Z'$) comme égale à $\mathcal{E}(Z - Z')^2$ puisque (ζZ) n'est plus ϑZ ; d'ailleurs on ignore quelle expression convient pour ($Z Z'$)



Conclusion.

Les informations supplémentaires données conduisent au partage de la perte d'information $p(Z)$ en deux composantes dont une irréductible : $p(Z^\circ)$. On peut les figurer sur deux axes sécants d'origine ζ substitués à l'axe unique. On peut considérer la figure

formée d'un axe unique ζZ^0 et d'axes transversaux $Z^0 Z Z'$ dont chacun correspond à un effectif donné de l'échantillon (il n'existe pas de seconde famille d'axes).

(ZZ') est un gain ou une perte d'information suivant les cas, mais ne signifie en général rien d'autre que $p(Z') - p(Z)$.

D - CAS OU L'ON A DES INFORMATIONS SUR LA LOI THEORIQUE DE DISTRIBUTION.

1/ - Généralités.

Jusqu'ici la théorie ne comportait aucun recours à des hypothèses sur une loi théorique (par exemple laplacienne) que les variables seraient censé suivre.

Cette ignorance de la loi théorique de distribution est propre à la théorie des sondages; elle se justifie du fait que les populations sondées (au cours des enquêtes sociales, économiques, culturelles, etc.) fournissent peu d'exemples de distributions statistiques remarquables.

Beaucoup de sondages sont occasionnels et on ignore (avant d'avoir fait l'enquête) qu'on va sonder par exemple une population distribuée suivant une loi de Galton. Les enquêtes périodiques, même, offrent peu de possibilités, pour la raison qu'un questionnaire permet de recueillir sur chaque unité (i) des données X_i, Y_i, Z_i, T_i , etc. dont le nombre dépasse parfois la centaine: si le montant X des revenus de l'unité statistique (le ménage) suit une loi de Galton, il y a peu de chances que Y, Z, T , etc. soient dans le même cas.

On évite donc de faire un plan de sondage conçu en vue d'une variable X privilégiée. Mais il est tout indiqué d'employer les informations qu'on peut avoir sur la structure de la variable X , à construire des estimateurs spéciaux pour les paramètres de la loi de distribution de X . On ne le fait guère: et d'abord parce que l'emploi d'un plan de sondage stratifié, à plusieurs degrés, etc. bien adapté rend les calculs difficiles et d'ailleurs peu rentables.

Toutefois Aitchison l'a fait (Réf. UTTING-COLE, 4ème partie) pour une distribution présumée être de Galton, en appliquant un résultat de Finney (Réf.) à un sondage de l'Institut de Statistique d'Oxford (sondage supposé bernoullien).

2/ - Cas où l'échantillon a un grand effectif.

En admettant le plan de sondage simplifié ainsi à l'extrême, il convient encore de distinguer le cas où l'échantillon est assez grand pour rendre valables les résultats asymptotiques bien connus de la théorie de l'estimation. (On reviendra au n° 3 sur les autres cas).

On sait qu'alors l'estimation Z° du maximum de vraisemblance est sans biais appréciable, donc sa variance est pratiquement une perte d'information; et dans le cas régulier de Cramer (Réf.) $1/n H$ (H défini au Ch.IV, II, § 9) est le minimum (asymptotiquement équivalent à $\mathcal{V}Z^\circ$) de toutes les variances d'estimateurs sans biais possibles $\mathcal{V}Z$ construits à partir de la loi de distribution.

Tout ce qui est dit (en A et B) sur Z° (1ère interprétation géométrique notamment) est valable; à ceci près qu'il ne s'agit plus d'une multiplicité linéaire Z ; que la corrélation $\sqrt{e}$ entre Z et Z° remplace $\sqrt{1-\rho^2}$ (où ρ désignait ρ_{xy}); que Z° , Z , $(Z - Z_0)$ ont des lois-limites de Laplace-Gauss, ce qui entraîne

$$\mathcal{E}(Z - Z_0)^2 = \mathcal{V}Z - \mathcal{V}Z^\circ$$

En particulier on peut avec Fortet (Réf.) distinguer trois cas :

1) Il existe pour ζ un estimateur sans biais

$$\bar{Z} = S z_i / n.$$

2) Il existe pour une fonction $\tau(\zeta)$ un estimateur sans biais

$$\bar{T} = S t_i / n.$$

3) Le cas général régulier, où existe l'estimateur Z° (faute de mieux).

Remarque : Le cas non régulier.

Dans une communication à l'Institut International de Statistique (1957), Alan Stuart signale des travaux récents⁽¹⁾ (1946 à 1952) concernant le cas non régulier, c'est-à-dire celui où, les conditions de régularité du 3ème cas n'étant pas remplies, l'estimation Z° du maximum de vraisemblance cesse d'exister ou perd ses droits. Il existe alors des cas où (l'estimation Z° n'existant pas) la variance de toute estimation ne peut descendre en-dessous d'une certaine limite qui est plus grande que $1/n H$ et qui peut jouer un rôle analogue à $\mathcal{V}Z^\circ$.

3/ - Cas où l'échantillon est de taille médiocre.

Les cas (1) et (2) ci-dessus restent inchangés. Mais au cas (3) $\mathcal{V}Z^\circ$ n'est plus une perte d'information, Z° étant biaisé. On peut lui

(1) Bhattacharyya	Sankya, 8.1, 1946.
Chapman & Robbins	Ann. Math. Stat. 22(581) 1951.
Kiefer	Ann. Math. Stat. 23(627) 1952.

faire correspondre entre autres la perte $\lambda H/n$ et plus généralement $\mathcal{E}W[Z]$ avec [cf. Ch. IV § II 7]

$$W = \Phi[Z] - \Phi[\zeta] + \sum \lambda_j (a_j - A_j)$$

$$\text{en posant } \mathcal{E}y_i = \sum_0^\infty a_j \zeta^j = 0, \quad S y_i/n = \sum_0^\infty A_j Z^j = 0$$

(On écrit Z au lieu de Z° , car il n'est plus vrai que $p(Z)$ soit minimum).

Posant de même $\Phi[\zeta] = f(a_0, a_1, a_2, \dots) = f$

$$\Phi[Z] = f(A_0, A_1, A_2, \dots) = f^*$$

et supposant la fonction f dérivable, on aurait

$$\frac{\partial W}{\partial A_j} = \frac{\partial f^*}{\partial A_j} - \lambda_j$$

qui, pour $Z = \zeta$ se réduit à $\lambda_j = \partial f / \partial a_j$

d'où

$$W = f^* - f + \sum \frac{\partial f}{\partial a_j} \cdot (a_j - A_j)$$

Toute fonction f convient alors, pourvu toutefois que la condition de convexité soit remplie :

$$\sum_j \sum_h \frac{\partial^2 f}{\partial a_j \partial a_h} (A_j - a_j) (A_h - a_h) \quad \text{définie négative.}$$

Toutefois il est exclu qu'on ait $\frac{\partial f}{\partial a_j} = \zeta^j$: W serait identique à 0.

Plus intéressant est le cas où f est elle-même une forme quadratique

$$f = \sum \sum b_{jh} a_j a_h$$

définie négative. D'où :

$$W = \sum \sum b_{jh} (A_h - a_h) A_j$$

On notera que $\bar{Y}(\mathcal{E}y_i) = \sum \sum \zeta^{j+h} A_j a_h \quad (= 0)$

$$\bar{Y}^2 = \sum \sum \zeta^{j+h} A_j A_h$$

d'où $\bar{Y}^2 = \sum \sum \zeta^{j+h} (A_h - a_h) A_j$

Lorsque les b_{jh} sont ζ^{j+h} , on retrouve donc

$$\mathcal{E} W = H/n,$$

qui correspond à

$$\frac{\partial f}{\partial a_j} = 0$$

(Voir Ch. IV, § II, 9).

Il serait beaucoup plus utile de connaître une expression de perte qui (pour n infini) serait asymptotiquement équivalente à $1/H n$ (comme l'est la variance de Z). Nous ne connaissons que

$$W = \bar{Y}^2/H^2$$

Des recherches ultérieures seraient nécessaires.

II - LE CHANGEMENT D'ECHANTILLON (à taille ou à coût constant) -

A - GENERALITES.

Après l'étude des suites de plans de sondage (d'une même famille) à effectifs décroissants, passons à la comparaison de plans de sondage (π) , (π') appartenant à des familles distinctes.

Ces comparaisons n'ont de sens qu'à effectifs constants⁽¹⁾ ou à coûts constants.

Avec la méthode classique des sondages, la variance d'échantillonnage est employée constamment pour comparer la qualité de 2 (et plus) estimations du même paramètre ζ ; il s'agira :

- d'un même estimateur, employé avec des méthodes différentes de tirage de l'échantillon; par exemple avec une urne unique, on comparera les variances relatives aux tirages avec et sans remise des boules tirées dans l'urne;

- d'estimateurs différents, appliqués (sans changer la structure de sondage) à des échantillons tirés au sort par des procédés différents; par exemple avec probabilités égales d'une part, probabilités inégales de l'autre;

(1) A supposer que ceci ait encore une signification.

- voire d'estimateurs différents ou non, appliqués à des structures de sondage différentes; par exemple quand on modifie le nombre de strates, les caractères stratificateurs, le contenu des unités de sondage, ... bref le "découpage" de l'univers sondé.

Les problèmes du premier type sont d'ailleurs assez particuliers, l'absence de biais

$$\mathcal{E} Z = \mathcal{E} Z' = \zeta$$

étant postulée.

Il serait commode (et satisfaisant pour l'esprit) de pouvoir parler du gain d'information relatif à l'emploi de Z au lieu de Z' pour désigner la différence (positive) $\mathcal{E}(Z' - Z)^2$.

Mais il n'y a en général aucune raison a priori pour qu'on ait le droit de parler ainsi, c'est-à-dire de traiter cet écart positif comme la variance :

$$\mathcal{E}(Z' - Z)^2$$

Bien entendu, il est exclu que les deux échantillons à comparer soient indépendants, auquel cas on aurait :

$$\mathcal{E}(Z' - Z)^2 = \mathcal{V} Z' + \mathcal{V} Z$$

au lieu de :

$$\mathcal{E}(Z' - Z)^2 = \mathcal{V} Z' - \mathcal{V} Z$$

En fait quand on écrit cette dernière relation, autrement dit :

$$\mathcal{E} [(Z' - Z)(Z - \zeta)] = 0 \quad (1)$$

on suppose l'existence d'un champ de probabilité sur lequel sont définis simultanément Z et Z', étant entendu que la loi de probabilité de Z (quel que soit Z') et celle de Z' (quel que soit Z) sont imposées. Le symbole $\mathcal{E}$ se trouvera du même coup défini.

Que signifie exactement (1), condition d'orthogonalité de $\overline{\zeta Z}$ et $\overline{ZZ'}$?

Sauf dans le cas de distributions particulières, il faut considérer que, une fois l'échantillon tiré avec le plan de sondage π , les $(Z - \zeta)$ sont des constantes quelconques et qu'il est nécessaire (et suffisant) d'avoir :

$$E(Z') = Z$$

où E désigne l'espérance mathématique lorsqu'est connu le grand échantillon (d'où on déduit la valeur Z).

On peut interpréter également (1) comme :

$$\mathcal{E}(Z' \mid Z) = Z$$

c'est-à-dire espérance mathématique liée par la connaissance de la valeur de Z ; ces points de vue ne sont pas tout à fait équivalents : quand on entre dans le domaine des grands échantillons (population très nombreuse elle-même, les tirages pouvant en outre être avec remise), il est bien naturel qu'on puisse retrouver plusieurs fois la même valeur de Z . Enfin Z et Z' peuvent avoir une loi-limite, l'échantillon devenant infiniment grand; auquel cas la formule est forcément :

$$\mathcal{E}(Z' \mid Z) = Z$$

A supposer l'existence de deux lois-limites de Laplace-Gauss, avec $\forall Z' > \forall Z$, l'ellipse de probabilité de la loi de (Z, Z') doit admettre la première bissectrice comme diamètre conjugué de la direction verticale.

En dehors de l'hypothèse $(n \rightarrow \infty)$, la distribution commune $(Z, Z')_n$ doit admettre comme courbe de régression de Z' en Z la droite d'équation $Z' = Z$; autrement dit, on doit avoir

$$Z' = Z + \vartheta, \quad \text{avec} \quad \mathcal{E}(\vartheta \cdot Z) = 0$$

(et $\mathcal{E}\vartheta = 0$, qui est vérifiée si Z et Z' sont tous deux estimateurs sans biais de ζ).

Sauf si n est grand, $(E Z' = Z)$ a la signification suivante :

Partant d'un échantillon du plan π (d'où une valeur de Z), on lui fait correspondre un ensemble d'échantillons du plan π' , d'où un ensemble de valeurs Z' . On s'arrange pour avoir :

$$E Z' = Z$$

On s'arrange pour que, si l'on opère ainsi à partir de tous les échantillons équiprobables de π , on obtienne (un même nombre de fois) tous les échantillons équiprobables de π' .

D'ailleurs il suffit qu'une telle opération soit possible (sans aller jusqu'à l'exécution) pour qu'on ait le droit de convenir que

$\mathcal{V}Z' - \mathcal{V}Z = \mathcal{E}(Z - Z')^2$. On va rencontrer des cas où l'opération n'est pas possible.

B - 1ère INTERPRETATION GEOMETRIQUE.

Z et Z' étant représentés chacun dans un espace auxiliaire, avec

$$\mathcal{V}Z = \overline{OM}^2 \quad \text{et} \quad \mathcal{V}Z' = \overline{OM'}^2$$

on pourra avoir :

$$\overline{MM'}^2 = \overline{OM'}^2 - \overline{OM}^2$$

si la représentation de Z' est projetable orthogonalement sur celle de Z.

En particulier s'il s'agit de représentations à une seule dimension, il est nécessaire que les divisions découpées par Z et Z' (sur leurs axes respectifs) soient semblables.

1/ - Quelques cas de représentation à une dimension (estimation d'une moyenne).

Deux divisions d'abscisses $\sigma^2(v - n/n(v - 1))$ et $\sigma'^2(v - n/n(v - 1))$ sont semblables et même homothétiques. Sont semblables également :

$$\frac{v\sigma^2}{v-1} \left(\frac{1}{n} - \frac{1}{v} \right) \quad \& \quad \frac{v'\sigma'^2}{v'-1} \left(\frac{1}{n} - \frac{1}{v'} \right)$$

(y compris pour v'_{∞} : sondages exhaustif et bernoullien)

$$\frac{k v \sigma^2}{k v - 1} \left(\frac{1}{kn} - \frac{1}{kv} \right) \quad \& \quad \frac{v \sigma'^2}{v-1} \left(\frac{1}{n} - \frac{1}{v} \right)$$

(sondages en grappes de même taille k).

etc.

2/ - Exemple : Tirage exhaustif et tirage bernoullien, avec $n = 2$.

Echantillon exhaustif a, b $\overline{X} = (a + b)/2$

Echantillon bernoullien (a, b) $\overline{X}' = (a + b)/2$
 (a, a) $\overline{X}' = (a + a)/2 = a$

Il y a une probabilité $v - 1/v$ que la seconde boule tirée diffère de la première. Il vient :

$$\mathcal{E}(\overline{X}' - \overline{X})^2 = \frac{v-1}{v} \cdot 0 + \frac{1}{v} \left(\frac{b-a}{2} \right)^2$$

Mais tous les couples (a, b) sont également possibles; d'où :

$$\begin{aligned}\mathcal{E}(\bar{X}' - \bar{X})^2 &= \frac{1}{v} \frac{\Sigma(b-a)^2}{2v(v+1)} = \frac{\sigma^2}{2(v-1)} = \frac{\sigma^2}{2} \left[1 - \frac{v-2}{v-1} \right] \\ &= v \bar{X}' - v \bar{X}\end{aligned}$$

3/ - Uncas de représentation à plusieurs dimensions (estimation d'une moyenne) :

Tirages exhaustif et bernoullien dans un sondage stratifié.

Si dans chaque strate l'échantillon bernoullien est déduit - d'un échantillon exhaustif donné - par un procédé aléatoire convenable, on a :

$$\mathcal{E}(\bar{X}_h' - \bar{X}_h)^2 = \mathcal{E}(\bar{X}_h' - \bar{x}_h)^2 - \mathcal{E}(\bar{X}_h - \bar{x}_h)^2$$

et, en pondérant par $(v_h/v)^2$ et totalisant, - en admettant en outre l'indépendance des opérations de chaque strate par rapport aux autres -, il vient :

$$\mathcal{E}(\bar{X}' - \bar{X})^2 = \mathcal{E}(\bar{X}' - \bar{x})^2 - \mathcal{E}(\bar{X} - \bar{x})^2$$

C - 2ème INTERPRETATION GEOMETRIQUE.

Z et Z' sont représentés dans un espace euclidien auxiliaire , avec

$$vZ = \overline{OM}, \quad vZ' = \overline{OM'}$$

Quand les pertes d'information sont de la forme :

$$vZ = \alpha + \beta + \gamma, \quad vZ' = \alpha' + \beta' + \gamma'$$

il vient :

$$\mathcal{E}Z' - vZ = (\alpha' - \alpha) + (\beta' - \beta) + (\gamma' - \gamma)$$

Il est clair que, si α' est fonction linéaire de α , de même β' et β , γ' et γ , - alors $vZ' - vZ$ est aussi une perte d'information, et c'est $\mathcal{E}(Z' - Z)^2$.

Exemple : Sondages à 2 degrés. Comparaison des tirages exhaustif et bernoullien. Cas particulier où $n_h = \bar{n}$; $v_h = \bar{v}$.

α' est proportionnel à $1/m$, et α à $(1/m - 1/\mu)$

β' est proportionnel à $1/n$, et β à $(1/n - 1/v)$
 γ' est proportionnel à $1/mn$, et γ à $(1/m - 1/\mu)(1/n - 1/v)$

On est donc bien dans le cas où $\mathcal{V}\bar{X}' - \mathcal{V}\bar{X}$ est une variance.

Remarque.

Il est clair que, si le mode de tirage différait pour le 2ème degré seul, ou pour le 1er degré seul, il suffirait de déduire l'échantillon bernoullien de l'échantillon exhaustif par le procédé aléatoire.

Plus généralement, partant d'une sous-population de $(m \times \bar{n})$ unités distinctes (dont la moyenne est $\bar{X}$), on éliminera (par tirages au sort équiprobables) un nombre quelconque λ d'unités qu'on remplacera (par tirages au sort équiprobables) par λ unités figurant déjà dans l'urne et qui y figureront ainsi deux fois), etc

Ainsi il est possible de construire un échantillon bernoullien à partir de l'échantillon exhaustif supposé donné, et de telle façon que :

$$\mathcal{E}(\bar{X}' - \bar{X})^2 = \mathcal{V}\bar{X}' - \mathcal{V}\bar{X}.$$

D - LE CAS GENERAL.

1/ - Il n'y a pas de distance entre deux estimateurs.

La comparaison des tirages bernoulliens et exhaustifs constitue en fait un cas d'exception; car lorsqu'on passe en revue les diverses circonstances courantes où la qualité d'estimateurs Z et Z' se juge en étudiant le signe de $\mathcal{V}Z - \mathcal{V}Z'$, on est forcé de constater que cette différence n'est ni $\pm \mathcal{E}(Z - Z')^2$, ni tout autre distance entre Z et Z' , ou entre les plans de sondage (π) et (π') . En effet (tout en ayant : $\mathcal{E}Z' = \mathcal{E}Z = \zeta$) : (π') est conçu de façon : soit à avoir $\mathcal{V}Z - \mathcal{V}Z' \geq 0$ (en tout état de cause, ou bien dans la majorité des cas); soit à réduire le coût d'enquête, avec une certitude ou des chances sérieuses d'avoir alors : $\mathcal{V}Z - \mathcal{V}Z' < 0$.

Dans le premier cas, on s'arrange pour que les valeurs de Z les plus éloignées de ζ ne soient pas (ou soient très rarement) prises par Z' ; pour cela certains échantillons E du plan (π) sont impossibles avec le plan (π') . Dans le second cas, on arrive au même fait pour une simple raison de commodité.

Ainsi à un certain champ de valeurs de Z ne correspond aucune valeur de Z' ; et parler de

$$\mathcal{V}(Z' - Z)^2 = \mathcal{V}Z + \mathcal{V}Z' - 2 \text{Cov}(Z, Z')$$

ne signifie rien, car $\text{Cov}(Z, Z')$ n'est pas défini; alors parler de la distance (Z, Z') risque fort de ne rien vouloir dire. On vise ainsi les cas suivants :

Comparaison d'un sondage systématique ou en grappe et du sondage ordinaire;

Comparaison d'un sondage stratifié et du sondage ordinaire;

Comparaison d'un sondage "équilibré" ("balanced sampling") et du sondage ordinaire;

Comparaison des tirages avec probabilités égales et probabilités inégales, etc.

Ajoutons que, si l'on compare par exemple deux découpages distincts des unités de sondage, ou deux découpages des strates, il peut ne plus même exister de champ où Z et Z' soient définies simultanément.

2/ - Que calcule-t-on en fait ?

Cependant la technique des sondages nous apprend, connaissant l'échantillon E' de (π') , à estimer non seulement $\varphi Z'$, mais aussi φZ et $\varphi Z - \varphi Z'$ ⁽¹⁾. On pourrait d'ailleurs estimer φZ et $\varphi Z'$ séparément avec deux enquêtes distinctes, avec respectivement un échantillon E de (π) et E' de (π') ; mais il serait à craindre que l'erreur d'échantillonnage sur $\varphi Z - \varphi Z'$ (estimations indépendantes) soit plus grande que $(\varphi Z - \varphi Z')$ elle-même; en employant un seul échantillon (soit E') on ne court guère de risque de se tromper sur le signe même de $\varphi Z - \varphi Z'$.

Faisons quelques remarques :

a) On estime Z' (et non Z) à l'aide de E' .

b) On estime φZ et $\varphi Z'$ à l'aide de E' ; mais on ne saurait estimer $\varphi Z'$ à l'aide de E (échantillon de π). Donc E' permet de calculer une sorte de distance entre le plan de sondage (π') et le plan de référence (π) , et non le contraire.

c) On sait (souvent)⁽²⁾ estimer sans biais φZ , mais il ne faut pas s'illusionner sur la signification de ces mots :

(1) Ceci fut pour nous un objet de profond étonnement à la première lecture du rapport de R. Jessen sur les sondages agricoles en IOWA (Réf. T1).

(2) Dans les cas les plus simples, l'analyse de variance fournit un mécanisme de calculs correct (du moins en première approximation).

Si l'on tirait (du même plan (π')) un nombre infini d'échantillons E' indépendants, on reconstituerait la distribution théorique de Z' et celle de $\mathcal{V}Z'$, avec la répartition statistique des valeurs des estimateurs. En revanche la distribution des valeurs calculées pour estimer $\mathcal{V}Z$, tout en admettant $\mathcal{V}Z$ pour espérance mathématique, ne coïnciderait pas du tout avec la distribution théorique de $\mathcal{V}Z$ (telle qu'on l'obtiendrait au contraire avec une infinité d'échantillons indépendants E du plan π).

d) On peut toujours écrire $\mathcal{E}(Z_1 - \zeta)^2 - \mathcal{E}(Z' - \zeta)^2 = \mathcal{E}(Z_1 - Z')^2$ avec $\mathcal{E}(Z_1 - Z')(Z' - \zeta) = 0$, c'est-à-dire $\mathcal{E}'[Z_1 - Z' \mid Z'] = 0$, où $\mathcal{E}'$ désigne l'espérance mathématique dans le champ de probabilités de Z' ; les valeurs de Z_1 (pour un échantillon donné de π') admettent Z' pour moyenne arithmétique; puis on s'arrange pour avoir $\mathcal{E}'(Z_1 - \zeta)^2 = \mathcal{V}Z$, sans pour autant que Z_1 et Z aient la même distribution.

On voit, en résumé, que (même en s'en tenant aux estimations sans biais) les écarts entre pertes d'information se prêtent mal à une interprétation simpliste comme distance entre deux points.

CHAPITRE VIII

SUR DIVERS CONCEPTS D'INFORMATION APPLICABLES AUX SONDAGES

I - LA QUANTITE D'INFORMATION DE R.A. FISHER -

Alors que nous voyons dans $1/n H$ la plus petite perte d'information qu'on puisse obtenir (à condition d'employer l'estimateur adéquat), R.A. Fisher appelle $n H$ l'information contenue dans l'échantillon, c'est-à-dire le plus grand gain d'information que puisse fournir le même échantillon (du moins si l'on est bien dans le "cas régulier") (Réf. Fi). On suppose toujours n très grand.

Dans le cas de l'estimation simultanée sans biais de deux paramètres ζ_1, ζ_2 d'une loi de distribution, la perte d'information $\mathcal{V}(u_1 Z_1 + u_2 Z_2)$ est une forme quadratique; égalée à W^2 c'est l'équation tangentielle d'une ellipse; lorsque n est assez grand, Dugué a établi l'existence de l'ellipse d'information intérieure à la précédente quels que soient Z_1 et Z_2 ; son équation est $Q = W^2$; la matrice de la forme quadratique Q est ce que Fisher a défini comme étant l'information.

Tels sont les rapports entre la perte d'information et l'information de Fisher.

II - EXTENSION AUX SONDAGES BERNOULLIENS -

Pour une loi de Laplace, l'estimateur efficace de $\bar{x} = \mu$ est la moyenne $\bar{X}$, la variance minimum des estimateurs est σ^2/n et l'information au sens de Fisher est n/σ^2 . Ceci s'étend à toutes les lois qui admettent un résumé exhaustif d'ordre 1.

L'idée la plus naturelle est donc de chercher à faire une théorie où l'information serait $1/\mathcal{V}\bar{X}$, c'est-à-dire n/σ^2 quelle que soit la loi de distribution. (n étant quelconque cette fois).

On serait même tenté d'appeler quantité d'information l'inverse ($1/\mathcal{V}Z$) de la variance, quel que soit l'estimateur Z .

Sil'on effectue les tirages bernoulliens dans une urne comprenant un nombre fini de boules, la définition précédente serait compatible avec la formule de récurrence (de Schutzenberger). Faisons deux tirages successifs d'une boule dans une urne renfermant quatre boules; la quantité d'information apportée serait :

$$\frac{1}{\sigma^2} + \frac{1}{4} \sum_1^4 \frac{1}{\sigma^2} = \frac{2}{\sigma^2}$$

et pour n tirages

$$n/\sigma^2 = 1/\overline{\psi X}$$

III - CAS DES SONDAGES EXHAUSTIFS -

En revanche avec des tirages exhaustifs, la même formule donne , pour deux tirages :

$$\frac{1}{\sigma^2} + \mathcal{E}\left(\frac{1}{s_1^2}\right)$$

où s_1^2 désigne la variance des ($v - 1$) boules restantes après le premier tirage.

De l'identité

$$\mathcal{E}\left(\frac{v-1}{v-2} s_1^2\right) = \frac{v}{v-1} \sigma^2$$

on ne peut tirer la valeur de $\mathcal{E}(1/s_1^2)$ mais on peut affirmer en tous cas qu'on n'a pas en général :

$$\mathcal{E}\left(\frac{1}{s_1^2}\right) = \frac{v}{v-2} \frac{1}{\sigma^2}$$

et par conséquent qu'on n'a pas :

$$\frac{1}{\sigma^2} + \mathcal{E}\left(\frac{1}{s_1^2}\right) = \frac{1}{\sigma^2} \left(1 + \frac{v}{v-2}\right) = \frac{2}{\sigma^2} \frac{v-1}{v-2} = \frac{1}{\overline{\psi X}}$$

En résumé : sil'on s'en tenait aux tirages indépendants, on pourrait accepter de transposer l'information de Fisher en théorie des sondages, de telle sorte que l'inverse de la variance de l'estimateur mesure de l'information apportée par celui-ci.

Mais on ne peut étendre valablement cette dernière convention si l'on passe du tirage bernoullien au tirage exhaustif des échantillons.

IV - DISCUSSION -

a) D'ailleurs il n'est pas rationnel de partir du sondage bernoullien et d'essayer de remonter au sondage exhaustif. Ce qu'il convient de faire, c'est au contraire de poser pour le sondage exhaustif des définitions cohérentes, puis d'en déduire celles du sondage bernoullien par passage à la limite lorsque l'effectif de la population tend vers l'infini.

b) En outre (mais ceci n'est pas un argument déterminant) le concept d'information ne correspond pas, quand on se place au point de vue des sondages, à celui posé par Fisher. Si je possède déjà un échantillon de 3 000 unités de sondage par exemple, et si j'en prends 100 de plus, il ne paraît guère possible d'admettre que ces 100 unités m'apportent autant d'information que les 100 premières unités qu'on a tirées de la population (ceci abstraction faite de toute considération de coût). C'est pourquoi, même lorsqu'on tire l'échantillon à la manière de Bernoulli, la définition de l'information inspirée par Fisher ne nous paraît pas finalement à retenir.

A cet égard Lindley (dont le concept d'information diffère du nôtre) a adopté la même position. L'information "apportée" par une unité est pour lui fonction concave du rang de tirage.

Ceci n'empêchera d'ailleurs pas de continuer, comme par le passé, à utiliser l'inverse du rapport des variances pour juger de la précision relative - ou de l'efficacité⁽¹⁾ comme on dit - de deux estimateurs, du moment qu'il s'agit de précision ou d'efficacité, - non de l'information.

V - SOLUTION -

On peut chercher à bâtir d'abord une théorie de l'information pour les sondages exhaustifs dans une urne unique. Cette information doit - avant tout - satisfaire à la condition générale de récurrence. Or le Lemme

 (1) A noter que Fisher (1922) appelle efficacité, non le rapport des inverses de variance mais le rapport des effectifs d'échantillon nécessaires pour obtenir la même variance; définitions qui ne coïncident que si la variance est de la forme a/n (Voir, Stuart, Réf. citée).

$$\mathcal{E} \left(\frac{n}{n-1} s_n^2 \right) = \frac{v}{v-1} \sigma^2$$

a semblé particulièrement commode pour l'application de la formule de récurrence; et ceci a conduit à rechercher à repérer le niveau d'information par une fonction de la forme :

$$t(v, n) = k(v, n) \sigma^2$$

l'effectif de la population étant v , celui de l'échantillon étant n , σ désignant l'écart-type de la variable x dans la population sondée.

On appellera gain ou perte d'information l'écart entre deux niveaux. Ce niveau d'information doit s'élever lorsque n va de 1 à $(v-1)$.

Avec un échantillon de taille nulle, on n'a pas d'information du tout.

Avec un échantillon de taille v , l'information est complète, totale.

Le problème essentiel est de savoir choisir pour ces deux cas extrêmes des repères convenables. Il est naturel de chercher à repérer par 0 l'absence d'information, par A ou l'infini l'information complète. On dira alors qu'on repère ou mesure l'information apportée par l'échantillon.

On va montrer que ce point de vue est indéfendable pour les enquêtes par sondage, avec $t(v, n)$. Au contraire, rien ne s'oppose à ce qu'on repère par $-\infty$ l'absence d'information. On dira alors qu'on repère ou mesure l'information perdue par le sondage.

VI - IMPOSSIBILITE DE MESURER L'INFORMATION APPORTEE PAR UN SONDAGE AVEC $t(v, n)$ -

Si le choix de l'information reste largement arbitraire quand on n'envisage que le sondage exhaustif simple, il ne faut pas oublier qu'on se propose d'aboutir à une théorie applicable à tous les plans de sondage communément employés par les techniciens.

Une autre difficulté est de savoir comment vont se combiner les informations fournies par chacune des urnes sondées (strates, sous-strates, unités non élémentaires). Cette combinaison est-elle de même nature qu'une addition (déduction faite des doubles emplois), ou bien qu'un produit ? Mais faire le produit revient à ajouter les logarithmes et tout dépend finalement d'une convention de langage : va-t-on appeler u ou $\log u$ l'information ?

a) Sondage à deux degrés.

Considérons un sondage à deux degrés, avec le cas limite du sondage en grappe (tirages exhaustifs avec probabilités égales) : si on possède une information partielle sur certaines unités primaires et une information complète sur certaines autres (grappes entières), on n'a pas une information complète sur la population.

Par conséquent, que les informations apportées par chaque unité primaire s'ajoutent ou qu'elles se multiplient les unes par les autres, - on peut dire a priori que l'information complète ne doit pas être mesurée par $+\infty$; elle doit être exprimée par un nombre fini. On va le supposer par la suite⁽¹⁾.

b) Addition des informations.

Lorsqu'on effectue des sondages indépendants dans une même population, il semble a priori que les informations apportées devraient s'ajouter - du moins d'après les idées communément admises sur l'information.

De sorte que, lorsqu'on ajoute l'information d'un sondage à celle d'un recensement, on devrait retrouver l'information du seul recensement (vu que le sondage n'apporte rien de plus). Ainsi, logiquement, l'information complète (celle procurée par le recensement) devrait s'exprimer par $+\infty$. (On observera qu'il en est de même si l'addition porte sur les logarithmes des informations et non sur les informations elles-mêmes).

En réalité ce second point de vue est logique sans être tout à fait décisif. On peut convenir en effet de limiter les recherches à un seul sondage simple exhaustif, - ce qui exclut toute possibilité d'ajouter à un sondage à 100% un autre sondage exhaustif à 5%. Lorsque l'urne est vide, l'opération s'arrête.

Ainsi il n'y a peut-être pas de contradiction flagrante entre (1) et (2), mais il y a tout de même là une restriction notable, une option inévitable entre deux concepts.

c) Considérons à présent un sondage stratifié.

S'il est une strate sur laquelle on ne possède aucune information, il est conforme aux habitudes de pensée des "sondeurs" de dire qu'on ne possède aucune information sur la population entière (mais seule-

(1) Tel n'est pas le cas pour l'information de Lindley (voir plus loin § IX).

ment sur les autres strates); c'est d'ailleurs un point sur lequel le statisticien dépourvu de formation sondage a une opinion différente.

Une conséquence fâcheuse en découle si l'on convient de repérer l'absence d'information par le niveau 0 (et non par le niveau $-\infty$) :

Si l'on veut conserver cette vue des techniciens, il ne faut pas que les informations apportées sur chaque strate s'ajoutent ou se combinent linéairement pour donner l'information relative à la population; tandis qu'il n'est pas exclu que les informations de strate se multiplient les unes par les autres (donc que leurs logarithmes s'additionnent).

Or il est facile de voir que c'est là une exigence incompatible avec notre hypothèse de travail : $t(v, n) = k(v, n)\sigma^2$ valable à l'intérieur des strates.

Si l'on pose (avec des poids convenables $\overline{w}$)

$$\mathcal{L}t = \sum_h \overline{w}_h(v_h, n_h) \mathcal{L}t(v_h, n_h)$$

ceci implique (quand on épuise toutes les strates) :

$$\max \mathcal{L}t = \sum_h \overline{w}_h(v_h, n_h) \left[\mathcal{L}k(v_h, n_h) + \mathcal{L}\sigma_h^2 \right]$$

Considérons alors un sondage à deux degrés. Si toutes les unités primaires sont sondées, le sondage à deux degrés devient sondage stratifié, la perte a la forme ci-dessus; si au contraire toutes les unités secondaires sont sondées dans l'échantillon d'unités primaires, on a un sondage en grappes dont la perte d'information est de la forme :

$$\mathcal{L}[k(\mu, m)\sigma_0^2] \quad \text{avec} \quad \sigma_0^2 = \frac{1}{\mu} \sum (\overline{x}_h - \overline{x})^2$$

et quand l'échantillon recouvre toute la population

$$\max \mathcal{L}t = \mathcal{L}k(\mu, \mu) + \mathcal{L}\sigma_0^2$$

Comme on a supposé que l'information restait bornée si la population entière est tirée, les deux expressions de $\max \mathcal{L}t$ sont incompatibles en général (puisque elles imposent une condition aux paramètres de la distribution).

d) Un autre fait est qu'on devrait pouvoir (toutes choses égales d'ailleurs) appeler sondage optimal celui qui apporte le maximum d'information. La théorie classique connaît déjà les répartitions d'échantillon optimales au sens de Neyman (telles que la variance

d'échantillonnage et le coût du sondage soient l'un minimum et l'autre constant). Ainsi il est souhaitable que ces deux points de vue soient conciliables (sans coïncider pour autant).

Admettons que le coût $\mathcal{C}$ soit combinaison linéaire des effectifs échantillon de strate.

Si l'on convenait que les informations de strate s'ajoutent, un cas particulièrement gênant serait celui où l'information de strate serait de la forme :

$$\lambda_h n_h$$

puisque le maximum d'information (pour un coût total $\mathcal{C} = \sum_h c_h n_h$ donné d'avance) s'obtiendrait en concentrant l'échantillon sur les strates dont le coefficient λ_h / c_h est le plus grand (contrairement à la technique admise).

Or on trouverait effectivement une expression de ce type pour l'information apportée dans des cas très simples. Posons :

$$t(v, n) = \frac{v\sigma^2}{v-1} f(v, n)$$

Reprenons une idée de M. Fonsagrive (après M. Chartier, Réf. Chartier 2). On considère deux strates d'effectif v formées par tirage au sort exhaustifs dans une urne d'effectif $2n$. On tire n boules de chaque strate.

Il est raisonnable d'admettre que l'information apportée est en moyenne la même que si $(2n)$ boules étaient tirées de l'urne primitive. D'où une relation :

$$\mathcal{E} \left[\frac{v}{v-1} (\sigma_1^2 + \sigma_2^2) f(v, n) \right] = \frac{2v}{2v-1} \sigma^2 f(2v, 2n)$$

Grâce au lemme fondamental, on a :

$$\mathcal{E} \sigma_1^2 = \mathcal{E} \sigma_2^2 = \frac{2(v-1)}{2v-1} \sigma^2$$

$$\text{D'où :} \quad 2f(v, n) = f(2v, 2n)$$

cas particulier de l'équation fonctionnelle

$$\rho f(v, n) = f(\rho v, \rho n)$$

En partageant en deux l'urne primitive, on est parvenu à l'équation avec $\rho = 2$; mais en la partageant en ρ parties égales, on obtient la même équation pour les diverses valeurs entières de ρ . La solution de cette équation compatible avec les autres conditions est justement λn .

Il est donc établi qu'on ne peut mesurer l'information apportée par un sondage avec l'expression $t(v, n) = \sigma^2 k(v, n)$.

La notion de quantité d'information apportée ayant fait l'objet de travaux importants à la limite du domaine des sondages, on va en passer en revue quelques-uns.

VII - L'INFORMATION DE SHANNON -

Le vocable "information" employé d'abord par Fisher a été repris par Shannon pour l'étude de problèmes de transmission. Good⁽¹⁾ notamment a pu exprimer cette "opinion raisonnable" qu'il n'y a pas grand'chose de commun entre l'information de Shannon et celle de Fisher.

C'est Schutzenberger⁽²⁾ qui, plus tard, a montré qu'elles entraient toutes deux dans le cadre d'une théorie générale de l'information.

Shannon appelait information d'un message la somme des informations des divers symboles qui le composent, - et information d'un symbole déterminé l'espérance mathématique ou plutôt la moyenne, changée de signe, du logarithme de la fréquence de son apparition.

Le problème pour Shannon était de faire occuper le moins de temps possible une ligne télégraphique par les messages; il fallait pour cela choisir pour les divers symboles possibles un code optimum tenant compte de la fréquence de chaque symbole.

La quantité d'information d'un message détermine finalement le temps pendant lequel on permet au message d'occuper la ligne quand on emploie ce code.

De leur côté, les statisticiens éprouvent le besoin de pouvoir définir une sorte de mesure (additive) de leurs travaux, en fonction de quoi ils répartiraient les moyens limités (personnel, matériel, crédits) dont ils disposent.

(1) Discussion à la suite de la communication de Barnard, Réf. Bar.

(2) Réf. Sch.

La "capacité" de la ligne a pour homologue en statistique la capacité du bureau; mais quel sera l'homologue de la quantité d'information ? Ceci dépend de la manière dont on va transposer en statistique les deux notions de message et de fréquence.

Cette transposition a été tentée à diverses reprises et notamment par Barnard⁽¹⁾.

VIII - LA QUANTITE D'INFORMATION DE BARNARD -

Barnard envisage une théorie suffisamment générale pour que le même symbole représente aussi bien un message (en théorie des transmissions), un problème (pour un calculateur électronique ou toute autre machine à calculer), ou une proposition (en statistique considérée comme un secteur de la logique).

Notons en passant que la notion de "quantité d'information" de Shannon a été effectivement adoptée par la cybernétique et est finalement revenue jusqu'au statisticien comme partie intégrante du "jargon" des spécialistes du matériel électronique.

Le "message" est donc transposé en une "proposition", et la "fréquence" devient une "probabilité", la probabilité que la proposition soit exacte. C'est ici que les difficultés surgissent.

Quel intérêt aurait-on à remplacer une probabilité p variant de 0 à 1, par une expression $(-\log p)$ variant de 0 à l'infini ?

Est-il plus intéressant d'avoir affaire à des expressions qui s'ajoutent, ou au contraire qui se multiplient les unes par les autres, quand on envisage simultanément plusieurs propositions disjointes ?

$$\log p + \log p' = \log pp'$$

Barnard arrive vite à la conclusion que $-\log p$ ne mérite pas d'être tenu pour la quantité d'information.

L'information serait une certaine grandeur (voire une certaine fonction) qu'on peut attacher à une proposition, au même titre qu'une probabilité; mais ce n'est certainement pas une simple fonction de la probabilité (ni $-\log p$, ni une autre).

Intéressons-nous plus spécialement à l'information en matière d'estimation statistique :

(1) Réf. G. Barnard m'a prié (1960) de préciser qu'il avait exposé (1950) une façon de concevoir l'information, mais en admettait volontiers d'autres.

Etant donné une loi de distribution $f(x, \zeta) dx$, où le paramètre ζ est à estimer, soit X_i une valeur réellement observée pour x ; l'information apportée par cette observation pourrait être :

$$- \log f(X_i, \zeta)$$

autrement une certaine fonction de ζ , et même une fonction de ce que Fisher appelle la vraisemblance (likelihood).

Pour Fisher (et Barnard) il n'est ni probabilité a posteriori ni loi de Bayes; une probabilité est toujours a priori, avant qu'on ait fait les expériences, les observations; une fois franchi le pas, il n'existe plus de probabilités, mais des vraisemblances (et des informations).

Et Barnard précise que Shannon, s'il était à sa place, considérerait ζ comme fixe, X_i comme variable, et prendrait la moyenne de $(-\log)$ sur l'ensemble des valeurs que X_i est susceptible de prendre.

IX - LA QUANTITE D'INFORMATION DE LINDLEY -

Quand on admet la loi de Bayes, et que par conséquent on ne conçoit pas de théorie de l'estimation sans une loi de probabilité a priori, on est conduit assez naturellement à définir la quantité d'information en fonction de la probabilité a posteriori, et non plus de la "vraisemblance".

Il est alors possible de conserver presque intégralement la définition de Shannon; on abandonne seulement le signe moins qui précède le logarithme, de façon à définir, pour chaque état de nos connaissances, un niveau d'information néгатif; quand on gagne de l'information, on s'élève vers la cote zéro. Nous suivons ici le point de vue de Lindley, le plus récent (Réf. Li)⁽¹⁾.

Soit $p(\zeta)$ la probabilité a priori et $p(\zeta|x)$ la probabilité a posteriori de ζ sachant que x est le résultat de ou des observations.

Les niveaux d'informations successifs sont, pour des tirages bernoulliens :

$$I_0 = \int p(\zeta) \log p(\zeta) d\zeta$$

$$I_1 = \int p(\zeta|x) \log p(\zeta|x) d\zeta$$

$$\bar{I}_1 = \int I_1 p(x) dx \quad (\text{information moyenne})$$

(1) Il ne diffère pas essentiellement de celui de ses prédécesseurs, McMillan, Blackwell, Bohnenblust-Shapley-Sherman, etc., de l'avis de Lindley.

où $p(x)$ désigne la loi de probabilité a priori de l'observation x .

Se plaçant au point de vue "séquentiel", on définit le niveau I_2 d'information après deux résultats d'observation x, x' , en substituant

$$p(\zeta | x, x') \quad \text{à} \quad p(\zeta | x)$$

dans la formule; on définit de même le niveau I_3 avec

$$p(\zeta | x, x', x''),$$

etc.

Le gain d'information apporté par les unités de sondage tirées aux rangs $n+1, n+2, \dots, m$ est égal à la différence du niveau $I_m - I_n$. Lindley montre que c'est une fonction "concave".

Comparaison avec la perte d'information.

Le caractère essentiel de la perte d'information est que le niveau I_∞ n'est pas rejeté à l'infini. Au contraire, dans des cas très simples, on constate que le niveau I_n de Lindley tend vers l'infini avec n ; par exemple lorsqu'on a (exemple du § X)

$$I_n = \frac{1}{2} \log(\lambda^2 + n) + \text{constante}$$

Par ailleurs le niveau I_0 serait (pour la perte d'information) rejeté à l'infini; alors qu'ici on part toujours d'un niveau fini. Ceci tient au fait qu'on ne part jamais d'une absence complète d'information, on se donne toujours a priori une distribution (qui peut en fait provenir d'un sondage antérieur).

X - ETUDE D'UN EXEMPLE DE LINDLEY -

Etudions avec Lindley une population infinie, distribuée suivant une loi de Laplace-Gauss de moyenne aléatoire et d'écart-type σ .

Considérons a priori ζ comme une autre variable de Laplace-Gauss, de moyenne μ (constante celle-la) et d'écart-type τ .

Supposons d'abord l'échantillon d'effectif $n = 1$, la probabilité élémentaire du couple aléatoire (x, ζ) est

$$(2\pi)^{-1} (\sigma\tau)^{-1} \exp \left\{ -\frac{1}{2} \left[\frac{(x - \zeta)^2}{\sigma^2} + \frac{(\zeta - \mu)^2}{\tau^2} \right] \right\} dx d\zeta$$

et l'exposant de l'exponentielle peut aussi s'écrire :

$$-\frac{1}{2} \left[\frac{(\mathbf{x} - \mu)^2}{\tau^2 + \sigma^2} + \frac{\tau^2 + \sigma^2}{\tau^2 \sigma^2} \left(\zeta - \frac{\sigma^2 \mu + \tau^2 \mathbf{x}}{\sigma^2 + \tau^2} \right)^2 \right]$$

ce qu'on interprète en disant que la loi de probabilité a priori de $\mathbf{x}$ est de Laplace-Gauss, de moyenne μ et d'écart-type $\sqrt{\tau^2 + \sigma^2}$; et que la loi de probabilité a posteriori de ζ est de même type, mais de moyenne : $\frac{\sigma^2 \mu + \tau^2 \mathbf{x}}{\sigma^2 + \tau^2}$ et d'écart-type $\frac{\sigma \tau}{\sqrt{\sigma^2 + \tau^2}}$. Il est utile ici de faire ressortir que :

THEOREME -

Si ζ a une distribution a posteriori de Laplace-Gauss, la quantité d'information est égale au logarithme (neperien) de l'inverse de son écart-type, diminué de $\log \sqrt{2\pi}$ (démonstration immédiate).

En conséquence, le gain d'information $I_1 - I_0$ est égal à :

$$\log \sqrt{\sigma^2 + \tau^2} - \log \sigma = \frac{1}{2} \log \left(1 + \frac{\tau^2}{\sigma^2} \right)$$

Passons de là à un échantillon de taille $n = 2$; il vient :

$$\frac{(x_1 - \zeta)^2}{\sigma^2} + \frac{(x_2 - \zeta)^2}{\sigma^2} + \frac{(\zeta - \mu)^2}{\tau^2} = \frac{2\tau^2 + \sigma^2}{\tau^2 \sigma^2} \left(\zeta - \frac{\tau^2(x_1 + x_2) + \sigma^2 \mu}{2\tau^2 + \sigma^2} \right)^2 + \dots$$

et le gain d'information est :

$$\frac{1}{2} \log (\sigma^2 + \tau^2 / \sigma^2 + 2\tau^2) = I_2 - I_1$$

Avec un échantillon de taille n , le gain d'information est :

par rapport au cas où $n = 0$

$$\begin{aligned} I_n - I_0 &= \frac{1}{2} \log \left(1 + \frac{\lambda^2}{n} \right) + \log n && \text{en posant } \lambda^2 = \frac{\sigma^2}{\tau^2} \\ &= \frac{1}{2} \log (\lambda^2 + n) \end{aligned}$$

par rapport à un échantillon de taille $n - 1$, il est :

$$I_n - I_{n-1} = \frac{1}{2} \log \left(1 + \frac{\lambda^2}{n} \right) - \frac{1}{2} \log \left(1 + \frac{\lambda^2}{n-1} \right) + \frac{1}{2} \log \left(\frac{n}{n-1} \right)$$

Quant à la loi a posteriori de ζ , c'est une loi de Laplace - Gauss

de moyenne	$\frac{\sigma^2 \mu + n \tau^2 \bar{X}}{\sigma^2 + n \tau^2}$	et d'écart-type	$\sqrt{\frac{\tau^2 \sigma^2}{\sigma^2 + n \tau^2}}$
ou de moyenne	$\frac{\lambda^2 \mu + n \bar{X}}{\lambda^2 + n}$	et d'écart-type	$\sqrt{\frac{\sigma^2}{\lambda^2 + n}}$

On peut comparer gain d'information et réduction de la perte d'information (variance), ou plutôt (s'il'on fait abstraction de $\log(n/n-1)$ et de $\frac{1}{2}$) comparer

$$\log \left(1 + \frac{\lambda^2}{n} \right) - \log \left(1 + \frac{\lambda^2}{n-1} \right) \text{ et } \left(\frac{\sigma^2}{n} - \frac{\sigma^2}{n-1} \right)$$

Lorsque n est grand (en supposant τ^2 du même ordre que σ^2), ces deux expressions sont des infiniment petits équivalents (au facteur τ^2 près).

Quant à la loi a posteriori de ζ , elle correspond assez bien à nos habitudes de pensée : il suffit d'imaginer qu'un sondage antérieur nous a procuré une moyenne μ et une variance (estimée) τ^2 et qu'on les prend comme paramètres de loi de probabilité a priori; alors on retrouve comme moyenne de la loi a posteriori l'estimateur habituel.

La difficulté est qu'il faut donc toujours commencer par faire une première estimation sans probabilité a priori, - ce qu'on déguise en l'appelant : choix des paramètres μ et τ .

En dehors de ce cas simple, de peu d'intérêt pratique, l'information de Lindley (comme l'estimation par la méthode de Bayes) diverge franchement de la présente théorie.

XI - POINT DE VUE SUR CETTE INFORMATION -

Le problème de l'estimation statistique peut être abordé de trois façons différentes :

1) En se donnant la loi de répartition de X et la probabilité a priori du paramètre à estimer ζ : comme Lindley ;

2) En se donnant seulement la loi de répartition de X , - la grandeur à estimer étant une constante inconnue non aléatoire : c'est le point de vue de R. A. Fisher;

3) En ne précisant même pas la loi de répartition de X , loi qui en pratique n'est pas connue et ne présente que des rapports éloignés avec les lois théoriques usuelles ou non : c'est le point de vue de la théorie des sondages.

Le propre de la théorie des sondages est de faire des estimations en ignorant la forme des lois de distribution des variables en jeu.

A fortiori est-il indispensable d'ignorer les lois de probabilité a priori des ζ .

A - Cependant, dans le cas de populations sondées très grandes (et c'est un cas limite aussi important en pratique qu'en théorie) les $\bar{X}$ échantillon devenant indépendants, la théorie des résumés exhaustifs d'ordre I intervient.

Si en outre les échantillons sont eux-mêmes très grands (cas très courant en pratique) les lois-limites de Laplace-Gauss et la considération de l'estimation du maximum de vraisemblance entrent en jeu.

Le problème de premier plan des sondages est l'estimation de la moyenne arithmétique $\bar{x}$ de la population, par celle $\bar{X}$ de l'échantillon.

Si on a par ailleurs des raisons de penser que la distribution est, disons de Galton-Mac Alister-Gibrat, on pourra employer au lieu de $\bar{X}$ un estimateur X' , tel que $\forall X' < \forall \bar{X}$ en particulier l'estimateur du maximum de vraisemblance X^0 , tel que

$$\forall X^0 = 1/nH$$

C'est là (on l'a dit) un cas exceptionnel; il est de pratique courante de réduire la perte d'information par ce qu'on connaît sur la structure de la distribution, la loi restant non spécifiée.

Exemples - Sondage stratifié.

$$X = x_{hi} = \bar{x}_h + \theta_{hi} \quad \text{avec} \quad \mathcal{E} \theta_h = 0$$

sans spécifier les lois centrées θ_h et connaissant les effectifs de strate v_h , on abaisse la variance dans le rapport :

$$\Sigma v_h \sigma_h^2 / \Sigma v_h \sigma_h^2 + \Sigma v_h (\bar{x}_h - \bar{x})^2$$

dans le cas où les fractions sondées sont les mêmes dans chaque strate.

Si en outre les écarts-types de strate σ_h sont connus, on pose :

$$X = \bar{x}_h + \sigma_h \vartheta_h \quad \text{avec} \quad \mathcal{E} \vartheta_h = 0, \quad \mathcal{V} \vartheta_h = 1;$$

sans spécifier autrement les ϑ_h , le sondage "à la Neyman" abaisse la variance dans le rapport :

$$(\sum v_h \sigma_h)^2 / \sum v_h \sigma_h^2$$

Autre exemple : Estimation par une droite de régression.

En postulant que X a une structure telle que :

$$X = Y + \vartheta \quad \text{ou} \quad X = Y + \sigma(y) \vartheta \quad \text{avec} \quad \mathcal{E} \vartheta = 0$$

on sait abaisser la variance dans le rapport $(1 - \rho^2)$ (ρ étant le coefficient de corrélation entre X et Y, et $\bar{y}$ étant supposé connu.

Dans chacun de ces cas, dès qu'on spécifie que les variables ϑ sont laplace-gaussiennes, ou simplement suivent des lois admettant la moyenne arithmétique comme résumé exhaustif, il devient impossible d'abaisser davantage la variance, du moins à effectif constant.

Bref il est utile de se souvenir qu'il existe une loi de distribution alors même qu'on en ignore tout.

B - De même il est permis de chercher à tirer parti d'une théorie de l'information qui suppose connues :

- la loi de distribution de X (à quelques ζ près);
- la loi de probabilité a priori des ζ .

Le tout est de savoir quel profit tirer du point de vue suivant lequel ζ cesse d'être une constante inconnue pour devenir une variable aléatoire.

Lorsqu'on cherche à appliquer le calcul des probabilités à l'analyse de données naturelles, il est bien normal qu'on ait quelque hésitation sur le choix du schéma d'urne à employer pour introduire le hasard dans le raisonnement. En revanche, lorsqu'on opère sur des données provenant de tirages au sort dont on a monté soi-même tous les mécanismes (et c'est le cas des méthodes modernes de sondage), on est en droit de douter de l'utilité et de la légitimité de l'introduction de variables aléatoires supplémentaires.

Au reste on a repris l'exemple du § X (du à Lindley) pour chercher à en tirer parti. Considérons une distribution laplace-gaussienne (μ, σ^2) de moyenne μ et d'écart-type σ , dont on extrait un échantillon

bernoullien d'effectif n ; sa moyenne $\bar{X}$ suit à priori la loi laplace-gaussienne $(\mu, \sigma^2/n)$. Si μ est inconnu (mais σ connu) et si l'on connaît $\bar{X}$ (c'est-à-dire un certain échantillon), il est assez naturel de considérer la variable aléatoire auxiliaire M de distribution $(\bar{X}, \sigma^2/n)$ comme représentant l'estimation du paramètre inconnu μ . Cette dernière distribution est une probabilité a posteriori (connaissant l'échantillon). La loi de Bayes donne inversement une loi de probabilité a priori de M (et la loi de probabilité a posteriori de l'échantillon pour toute valeur donnée de M).

Pour alléger le calcul, on s'assurera de l'identité de deux lois gaussiennes à un nombre quelconque de variables en identifiant :

- d'une part les exposants de l'exponentielle (au facteur $-\frac{1}{2}$ près);

- d'autre part l'autre facteur de l'élément différentiel (à un multiple près de $1/\sqrt{2\pi}$).

La loi de Bayes s'écrit dans le cas d'un seul tirage :

$$p(x_1) \cdot p(M | x_1) = p(M) \cdot p(x_1 | M)$$

Identification : $\frac{(x_1 - \mu)^2}{\sigma^2} + \frac{(x_1 - M)^2}{\sigma^2} = \frac{(M - \mu)^2}{2\sigma^2} + \frac{2}{\sigma^2} \left(x_1 - \frac{M + \mu}{2} \right)^2$
ce qui fait bien apparaître la loi a priori $(\mu, 2\sigma^2)$ pour M .

Dans le cas de deux tirages, on a :

$$p(M) \cdot p(x_1, x_2 | M) = p(x_1, x_2) \cdot p(M | x_1, x_2)$$

avec $p(x_1, x_2) = p(x_1) \cdot p(x_2) = p(x_1 - x_2) \cdot p(x_1 + x_2)$

correspond aux égalités suivantes d'exposants :

$$(x_1 - \mu)^2 + (x_2 - \mu)^2 + \left(M - \frac{x_1 + x_2}{2} \right)^2 \cdot 2 = \frac{3}{2} (x_1^2 + x_2^2) + x_1 x_2 + 2(M^2 + \mu^2) - 2(x_1 + x_2)(M + \mu) = \left[(x_1 + x_2 - M - \mu)^2 + \frac{1}{2} (x_1 - x_2)^2 \right] + (M - \mu)^2$$

où le crochet correspond à $p(x_1, x_2 | M)$ et $(M - \mu)^2$ à $p(M)$.

Dans le cas de n tirages, il vient :

$$S(x_i - \mu)^2 + n(M - \bar{X})^2 = \left[S\left(x_i - \frac{M + \mu}{2}\right)^2 + n\left(\bar{X} - \frac{M + \mu}{2}\right)^2 \right] + (M - \mu)^2 \frac{n}{2}$$

avec la même décomposition en $p(x_1, x_2, \dots, x_n | M) \cdot p(M)$.

Conclusion.

La loi de Bayes conduit pour M à une loi de probabilités a priori (avant tirage) : laplace-gaussienne $\left(\mu, \frac{2\sigma^2}{n}\right)$.

Il en résulte que cette loi a priori ne serait pas la même quand on tire 1, 2, ... n boules, ce qui est absurde. Autrement dit :

Il n'est pas possible de représenter l'estimation (séquentielle) de μ par une variable aléatoire gaussienne M , constamment centrée en $\bar{X}$ et de variance σ^2/n , lorsqu'on suppose les variables X de valeur centrale μ inconnue mais fixe (comme c'est le cas lorsqu'on sonde une urne); on en conclut qu'il n'y a pas lieu de conserver l'aléatoire M .

Ceci ne signifie pas que, dans aucun cas, la loi de Bayes ne puisse avoir d'application. Mais nous la réserverions à des cas où il n'y a pas eu tirage au sort effectif (échantillon formé des premiers arrivants, par exemple).

XII - SUR L'INFORMATION SELON SCHÜTZENBERGER -

Schützenberger a établi (dans sa thèse, Pub. Inst. Stat. Uni. Paris, III, 1-2, page 43) le théorème suivant :

"Toute information⁽¹⁾ est la valeur moyenne, étendue à l'ensemble des états, de la résultante de l'application d'un opérateur linéaire S sur le logarithme de la probabilité a priori de chaque état".

En particulier avec $S = -1$, on a l'information de Shannon;

avec $S = -\frac{\partial^2}{\partial \zeta^2}$, on retrouve l'information de Fisher,
etc.

D'où la question : à quel opérateur linéaire S correspondrait la "perte d'information" et en particulier la variance ?

1/ Tout d'abord les "états" évoqués dans une théorie volontairement très générale peuvent être compris de plusieurs manières.

(1) On notera qu'il existe de nombreux cas de pseudo-information, c'est-à-dire d'expressions qui, sans être des informations, sont de la forme :

$$p_i S \log p_i$$

Limitons-nous d'abord au sondage exhaustif dans une urne unique. A une taille n donnée de l'échantillon correspond :

- un plan de sondage π ;
- $C_\nu^n = m$ échantillons distincts, donc m valeurs d'un estimateur Z ;
- m^2 couples d'échantillons distincts, donc m^2 valeurs $\frac{1}{2} (Z_i - Z_j)^2$ ou plus généralement (ij) , avec

$$(ii) = 0; \quad (ij) = (ji)$$

etc.

a) Un "état" peut donc signifier un échantillon, un couple d'échantillons, etc., l'information (perdue) étant celle du plan de sondage tout entier.

En pareil cas, les probabilités p_i sont égales entre elles, - à $1/m$ pour la première interprétation, à $1/m^2$ pour la seconde, etc.

Si par "état" on entend "couple", peut-on considérer $\frac{1}{2} (Z_i - Z_j)^2$, - plus généralement (ij) , - comme un $S \log(p_{ij})$?

b) Mais par "état" on peut entendre encore une chaîne d'échantillons de taille $(\nu - 1)$, $(\nu - 2)$, ... $(n + 2)$, $(n + 1)$, n , telle qu'on perde à chaque étape une unité de sondage de l'échantillon. Si les probabilités sont encore égales entre elles, leur expression est devenue :

$$n!/\nu!$$

Au lieu d'une seule chaîne, on peut d'ailleurs en considérer deux simultanément. Mais, de toute manière nos "états" sont équiprobables.

L'information sera obtenue en divisant la somme des $(S \log p_i)$ par le nombre d'états.

2/ Que faut-il comprendre alors par résultante de l'application d'un opérateur linéaire S ?

Considérons deux plans de sondage de tailles n et n' ; soit π et π' .

Supposons que, n étant plus grand que n' , on passe d'un "état" de π à un état de π' , donc d'une probabilité p à une probabilité $p' = pq$.

Par exemple :

$$p = n! / v! , \quad p' = n'! / v! , \quad q = n'! / n!$$

d'où :

$$\log p = \log n! - \log v! , \quad \log p' = \log n'! - \log v! ,$$

$$\log q = \log n'! - \log n!$$

a) On voit alors que $S \log p_i$ ne peut désigner $(Z - \zeta)^2$, pour la raison qu'on n'a pas :

$$(Z' - Z)^2 + (Z - \zeta)^2 = (Z' - \zeta)^2$$

l'égalité des deux membres supposant $(Z' - Z) \cdot (Z - \zeta) = 0$.

Il en résulte que "état" ne peut signifier "échantillon" avec lequel on calculerait Z (ou Z').

b) Désignons au contraire par état un couple (E_i, E_j) d'échantillons, chacun de taille n , et passons de là (en perdant $n' - n$ unités pour chacun) à un couple (E'_i, E'_j) chacun de taille n' .

On voit cette fois qu'on peut poser :

$$2S = (Z'_i - Z'_j)^2 - (Z_i - Z_j)^2$$

expression dont le second terme se réduit à 0 pour $Z = \zeta$ (c'est-à-dire $\pi = U$).

En résumé : Par état on peut comprendre le couple d'échantillons; et par résultante de l'application de l'opérateur linéaire S sur $\log p$, on peut entendre :

$$[(Z'_i - Z'_j)^2 - (Z_i - Z_j)^2] / 2$$

ou (plus généralement) toute fonction $G(I', J') - G(I, J)$ telle que $G(J, I) = G(I, J)$, $G(I, I) = 0$.

Généralisation.

Pour les besoins par exemple de la mesure de la perte d'information sur l'estimation d'une perte d'information, on pourra considérer comme état un groupe de 4 échantillons et pour

$$S \log p = (Z'_i - Z'_j)^2 (Z'_h - Z'_k)^2 - (Z_i - Z_j)^2 (Z_h - Z_k)^2$$

ainsi que des expressions plus générales.

Cas de plans de sondage plus complexes.

La transposition est immédiate. En particulier les tirages avec probabilités inégales (associés à des estimateurs sans biais) modifient le calcul de l'espérance mathématique de $S \log p$, mais non $S \log p$.

XIII - AUTRE INTERPRETATION DE LA PERTE D'INFORMATION (T. CR 4 1958) -

Revenons sur le théorème de Schützenberger évoqué au § XII.

Soit i, j , deux échantillons au hasard (distincts ou non) du même plan de sondage et F une fonction positive, telle que :

$$F(i, j) = F(j, i), \quad F(ii) = 0$$

La probabilité d'obtenir le couple ij est $1/m^2$ s'il existe m échantillons équiprobables pour ce plan de sondage.

Appelons écart $D(ij)$ entre les deux échantillons ij l'expression :

$$D(ij) = \left[\exp [- F(ij)] m^2 \right]^{-1}$$

et densité le rapport $m^{-2}/D(ij) = p(ij)$

$$\text{soit} \quad q(ij) = 1 - \exp [- F(ij)]$$

C'est bien une probabilité (comprise entre 0 et 1); posons de même :

$$p(ij) = \frac{d}{dF} q(ij) = \exp [- F(ij)]$$

$$\text{d'où} \quad - \log p(ij) = F(ij).$$

Appelons enfin $\mathcal{E} [- \log p(ij)] = \mathcal{E} [F(ij)] = \mathcal{E}(p_i)$ la perte d'information. On retrouve à la fois la définition de Shannon et la nôtre.

Donc, tout se passerait comme si la densité de probabilité de l'écart entre les deux échantillons ij était de la forme :

$$A \cdot \exp [- F(ij) \cdot B]$$

où A et B désignent deux constantes (positives).

En particulier : Cas de la variance.

Pour estimer sans biais le paramètre ζ , on dispose de

$$Z(i) \quad \text{et} \quad Z(j)$$

c'est-à-dire deux estimations, sur échantillons tirés au sort (i) et (j), avec l'estimateur sans biais Z .

$$\text{Soit} \quad F(ij) = [Z(i) - Z(j)]^2/2$$

On sait que : $(p_i) = \nabla Z = \mathcal{E}F$

Pour retrouver l'information de Shannon, il faut donc que $Z(i) - Z(j)$ ait une densité de probabilité de Laplace-Gauss.

Lorsque l'échantillon est grand, les moyennes sur échantillon admettent des lois-limite de Laplace-Gauss; de même les estimations du maximum de vraisemblance; etc. Supposons que ce soit le cas pour Z . La différence $Z(i) - Z(j)$ suit elle-même une loi de cette nature.

Ainsi, lorsque l'estimateur admet une loi-limite de Laplace-Gauss, il y a coïncidence entre l'information de Shannon et la perte d'information dans le cas des grands échantillons.

Mais si l'on écarte cette interprétation asymptotique, on est réduit à supposer que le plan de sondage est bernoullien et la distribution de base laplace-gaussienne, ce qui ne présente plus beaucoup d'intérêt pratique (§ X ci-dessus).

XIV - SUR L'INFORMATION SELON FISHER -

Finalement l'opérateur linéaire S de Schützenberger qui est utilisé en sondage, n'a donc rien de commun avec celui $(\partial^2/\partial \zeta^2)$ de Fisher.

La relation simple entre σ^2/n et n/σ^2 (avec une loi de Laplace-Gauss) apparaît comme accidentelle. Essayons de l'expliquer.

Limitons-nous au cas de l'estimation d'un seul paramètre ζ , par Z .

Soit F la loi de probabilité de Z , et G celle des variables $X_1, X_2, \dots, X_n$ indépendantes au début et qu'on lie par une valeur donnée à Z . On désigne par f_i la loi de probabilité de X_i . Il vient :

$$\sum_{(n)} \log f_i = \log F + \log G$$

En dérivant deux fois par rapport à ζ , en prenant les espérances mathématiques des deux membres et en changeant les signes, il vient :

$$\begin{aligned} \text{Information de Fisher} &= \mathcal{E} \left[- \frac{\partial^2}{\partial \zeta^2} (\log F) \right] + \mathcal{E} \left[- \frac{\partial^2}{\partial \zeta^2} (\log G) \right] \\ &= \mathcal{E} \left[\frac{\partial}{\partial \zeta} (\log F) \right]^2 + \mathcal{E} \left[\frac{\partial}{\partial \zeta} (\log G) \right]^2 \end{aligned}$$

Lorsque la loi $f(X)$ admet un résumé exhaustif d'ordre 1 par rapport à ζ , le deuxième terme du deuxième membre disparaît; il subsiste autrement.

Lorsque n est grand, $\bar{X}$ admet une loi-limite de Laplace-Gauss, et on va s'en tenir aux estimateurs Z possédant cette même propriété.

Par conséquent on aura :

$$\mathcal{E} \left[- \frac{\partial^2}{\partial \zeta^2} \log F \right] = \frac{n}{\sigma^2} + \epsilon$$

où σ n'est pas l'écart-type des X mais celui des Z pour $n = 1$ (ou à défaut pour $n = r$).

Considérons alors un couple d'échantillons de taille n , soit (1), (2).

Si $Z(1)$ et $Z(2)$ sont sans biais, la variable $Z(1) - Z(2)$ est indépendante de ζ . De plus, comme n est grand, elle est laplace-Gaussienne, comme la variable $Z(1) + Z(2)$ dont elle est indépendante.

Le moment du 2ème ordre de $Z(1) - Z(2)$, en particulier, ne peut dépendre de ζ . Au contraire, l'information de Fisher est une fonction de ζ , constante dans le cas particulier où $\log f$ est un polynôme du 2ème degré en ζ (exclusivement).

Pour le couple (1) (2), l'information de Fisher est :

$$\mathcal{E} \left[- \frac{\partial^2}{\partial \zeta^2} \log F(1 + 2) \right] + \mathcal{E} \left[- \frac{\partial^2}{\partial \zeta^2} \log G(1) - \frac{\partial^2}{\partial \zeta^2} \log G(2) \right]$$

où le premier terme est voisin de $2n/\sigma^2$. On sait que σ^2 ne dépend pas de ζ ; c'est donc dans le second terme qu'est localisée l'influence de ζ .

Finalement c'est cette présence de

$$\mathcal{E}(-\partial^2 \log G / \partial \zeta^2) = \mathcal{E}(\partial \log G / \partial \zeta)^2$$

à côté de $\mathcal{E}(\partial \log F / \partial \zeta)^2$ qui rend l'information de Fisher (pour n grand) au moins égale à l'inverse de $\mathcal{V}(Z)$.

XV - DERNIERES REMARQUES -

Les informations de Lindley ou de Fisher sont bien entendu des solutions de l'équation de Réurrence de Schützenberger. Mais l'espérance mathématique qui y figure a, pour l'information de Lindley, une signification fort différente de la nôtre; alors que nous lui donnons le même sens que Fisher. Le raisonnement que nous avons fait, à la IIIème partie du Chapitre IV n'est valable, il est vrai, que pour des pertes d'information finies; mais il serait possible de l'adapter au cas présent. Ainsi est-on conduit à poser :

$$\vartheta(\pi') - \vartheta(\pi) = n$$

$$1/H = E g(i, j) = \gamma$$

avec deux sondages bernoulliens (π') et (π) , d'effectifs :

$$\lambda \text{ et } \lambda + n$$

avec
$$2g(i, j) = \left[\frac{\partial}{\partial \zeta} \log \frac{f(x_i, \zeta)}{f(x_j, \zeta)} \right]^{-2}$$

Ainsi s'achève cet exposé qui, à notre regret, malgré sa longueur, laisse encore bien des points dans l'ombre.

BIBLIOGRAPHIE

Reference

- A.1. ANSCOMBE - Sequential Estimation. J. Royal Stat. Society - Series B. 1953.1.
- BAR G. A. BARNARD - The theory of information - J. Royal Stat. Society - Series B. 1951, p. 46.
- BA 1. D. BASU - On symmetric estimators in point estimation with convex weight functions.-SANKHYA 12, p.45 (décembre 1952).
- BA 2. D. BASU - On the optimum character of some estimators used in multistage sampling problems - SANKHYA, 13, p.3 63 (juin 1954).
- CHA 1. CHARTIER F. - Note sur les fluctuations aléatoires d'une variance estimée - Bulletin d'information de l'INSEE - n°3 mars 1957 - p. 31-40.
- CHA 2. CHARTIER F. - Sur l'influence d'une stratification aléatoire lors de l'estimation du total d'une population - Journal de la Société de Statistique de Paris 97, 4-5-6. 1956, p.121-129 .
- CR.1. Harald CRAMER - Mathematical Methods of Statistics - Princeton, 1946.
- CR.2. Harald CRAMER - Sur un nouveau théorème-limite de la théorie des probabilités - 1ère partie du fascicule : Les sommes et les fonctions de variables aléatoires - Actualités Scientifiques n°736; Hermann 1938.
- DA.1. G. DARMOIS - Sur les lois de probabilités à estimation exhaustive CR. 200 1935 - p.1265.
- DA.2. G. DARMOIS - Méthodes d'estimation - Actualités scientifiques Hermann 1936.
- DA.3. G. DARMOIS - Sur l'estimation des grandeurs par leurs mesures (Annuaire du Bureau des Longitudes. 1952).

Référence

- DU. D. DUGUE - Application des propriétés de la limite au sens du calcul des probabilités à l'étude de diverses questions d'estimation. - Journal de l'Ec. Polytechnique 1937 - p. 305.
- FE FERON - Information, Régression, Corrélation - Thèse pour le doctorat ès Sciences, Paris 1954.
- FIN FINNEY - On the distribution of a variate whose logarithm is normally distributed, page 155. Suppl^t of the J. of the Royal Stat. Society VII n°2, 1941.
- FIS 2 FISHER (Sir Ronald) - On the mathematical foundations of theoretical Statistics - Phil. Trans. Roy. Soc. London - Series A. 222 p. 309 - (1922).
- FIS 1 FISHER (Sir Ronald) - Theory of statistical estimation - Proceedings Cambridge Phil. Society 22 (1925) p. 700.
- FIS 3 FISHER (Sir Ronald) - Statistical methods and Scientific Inference - Oliver & Boyd. - 1956 - p. 148, etc.
- FOR R. FORTET - Calcul des probabilités - CNRS 1950.
- GE GEISSER (Seymour) - A note on the normal distribution. - The Annals of Math. Statistics - septembre 1956 - p. 858 - (Geary. J.R. Stat. Soc. Suppl. Vol. 3 (1936) n°2. (Réf.:) Lukacs. Ann. Math. Stat. Vol. 13 (1942) p. 91. (Daly. Ann. Math. Stat. Vol. 17 (1946) p. 71).
- KH Abbas. G. KHADJENOURI - Contribution à la théorie mathématique de l'échantillonnage - Thèse pour le Doctorat ès Sciences - PARIS 1956 -
- LI 1. D. V. LINDLEY - Statistical Inference. - J. of the Royal Stat. Society - 1953 - Séries B. XV n°1 - p. 30.
- LI 2. D. V. LINDLEY - On a measure of the Information provided by an experiment - The Annals of Math. Statistics - 27 n°4 déc. 1956 - p. 986.
- LI 3. D. V. LINDLEY - Binomial sampling schemas and the concept of information - Biometrika Vol. 44 n°1-2 - p. 1957.
- N.1. Jerzy NEYMAN - On the two different Aspects of the Representative method ; the method of Stratified Sampling and the method of Purposive Sampling. - Journal of the Royal Statistical Society 1934 - p. 558-625.
- N.2. Jerzy NEYMAN - Contribution to the Theory of Sampling Human Population - J. Amer. Stat. Assoc. 33 (1938) p. 101-116.
- N.3 Jerzy NEYMAN - Sur une famille de tests asymptotiques des

Référence

- hypothèses statistiques composées - Madrid - Revista Trabajos de Estadística - V. II. 1954.
- SC. M. P. SCHÜTZENBERGER - Contribution aux applications statistiques de la théorie de l'information (Thèse pour le doctorat ès-Sciences-Paris 1954) Publications de l'Inst. de Stat. de l'Université de Paris - III. 1. 2. 1954.
- ST. STUART (Alan) - The measurement of estimation and test efficiency - Session 1957 de l'Inst. Inst. Stat. - papier n°62.
- TU. TUKEY - Variances of Variance component - Ann. Math. St. 1956, p. 722.
- UT. J.E.G. UTTING & Dorothy COLE - Sample Surveys for the Social Accounts of the Household Sector - Bulletin of the Oxford University Institute of Statistics - janv. 1953.
- WAL. Abraham WALD - Statistical Decision Functions - New York 1950.
- Z L. ŽARKOVIĆ - Sampling Control of Literacy Data. Journal of the Am. Stat. Assoc. 1954 - pp. 510-519.
- T.CR Pierre THIONET - Notes aux comptes-rendus Acad. Sci. :
1. - Sur la variance de l'estimation d'une variance (Vol. 245, 1957, p. 2168).
 2. - Une généralisation de la variance d'échantillonnage dans le cas de tirages exhaustifs d'une urne (Vol. 245, 1957, p. 2464).
 3. - Représentation topologique des sondages (Vol. 246, 1958 p. 46).
 4. - Sur les rapports entre divers concepts d'information (Vol. 246, 1958, p. 123).
 5. - Sur les pertes d'information qui sont des fonctions de risque (Vol. 246, 1958, p. 128).
 6. - Sur les pertes d'information imputables à certaines estimations biaisées (Vol. 246, 1958, p. 536).
 7. - Sur une théorie générale des pertes d'information par sondage (Vol. 246, 1958, p. 692).
- T.IIS Pierre THIONET - Erreurs d'échantillonnage - fonctions de risque et information perdue - Session 1957 de l'Inst. Int. de Stat. - Papier n°64 - Bulletin d'Inf. de l'INSEE, sept-nov. 1957 - p. 51-64.
- T.L. Pierre THIONET - Méthodes statistiques modernes des administrations fédérales aux Etats-Unis. - Actualités Scientifiques. Hermann 1946.

Référence

- τ.2. Pierre THIONET - L'école moderne de statisticiens italiens - Journal de la Société de Statistique de Paris (1945 n° 11-12 1946 n° 1-2).
- τ.3. Pierre THIONET - La théorie des sondages - Etude théorique n° 5 de l'INSEE - Paris (I. N) 1953.
- τ.4. Pierre THIONET - Application des méthodes de sondage aux enquêtes statistiques - Etude théorique n° 6 de l'INSEE - Paris (I. N.) 1953).
- τ.5. Pierre THIONET - Théorie des sondages - Cours de l'Ecole d'application de l'INSEE (année scolaire 1955-56).
- τ.6. Pierre THIONET - En s'exerçant à la théorie des sondages - Bulletin d'information de l'INSEE - sept-octobre 1954.
- τ.7. Pierre THIONET - Recherche d'une fonction de coût applicable aux enquêtes par sondage de l'INSEE. - Revue de l'Institut International de Statistique 1954 - p. 33.
- τ.8. Pierre THIONET - Décisions à propos de sondages - Revue de Statistique appliquée - 1955 p. 71-80.
- τ.9. Pierre THIONET - Le problème théorique du plan d'échantillonnage - Journal de la Société de Statistique de Paris - 1948 - mars-avril.
- τ.10. Pierre THIONET - a) Considérations théoriques à propos de sondage (1955).
b) Notion d'Information et théorie des sondages (1956).
(Documents intérieurs au S. E. E. F.).
- τ.11. Pierre THIONET - Théorie des décisions statistiques et théorie des sondages.
(Bull. d'Inf. de l'INSEE - sept-oct. 1956, p. 1. 6).

TABLE DES MATIÈRES

	Page
INTRODUCTION	269
CHAPITRE I - Tirages équiprobables de boules dans une urne	273
CHAPITRE II - Les pertes d'information dans le cas de l'urne de Bernoulli	291
CHAPITRE III - La perte d'information avec des plans de son- dage quelconques.....	307
I - Définitions fondamentales	307
II - Topologie des sondages	316
CHAPITRE IV - La forme analytique des pertes d'information	327
I - Pertes au sens du chapitre I	327
II - Cas des estimations biaisées	336
III - Pertes au sens du chapitre II	357
IV - Pertes d'information d'un sondage ou d'une estimation	364
CHAPITRE V - Les sondages à plusieurs degrés	375
CHAPITRE VI - Sondages divers (notamment à deux phases) ..	395
CHAPITRE VII - Les informations supplémentaires	403
I - Le changement d'estimateur	403
II - Le changement d'échantillon	419
CHAPITRE VIII - Sur divers concepts d'information applicables aux sondages	427
BIBLIOGRAPHIE	451

DEUXIEME THESE

Propositions données par la Faculté

EXISTENCE ET UNICITE DES SOLUTIONS DES EQUATIONS
DE LA PRODUCTION ET DES ECHANGES, EN
ECONOMIE CONCURRENTIELLE.

Vu et approuvé
Paris, le 11 juin 1958
Le Doyen de la Faculté des Sciences,
J. PERES

Vu
et Permis d'imprimer :
Le Recteur de l'Académie
de Paris,
Jean SARRAILH

IMP. LOUIS-JEAN — GAP

Dépôt légal n° 229-1960

